REPORT RESUMES

ED 012 794 08 AA 000 154

FEASIBILITY OF PROVIDING TRADE COMPETENCY EXAMINATIONS FOR TEACHERS ON A NATIONAL BASIS.

BY- GRIESS, JERALD

NEW YORK STATE EDUCATION DEPT., ALBANY

REPORT NUMBER BR-5-0043-4 PUB DATE FEB 67
CONTRACT OEC-5-85-110
EDRS PRICE MF-$0.50 HC-$3.40 85P.

DESCRIPTORS- *SEMINARS, *TEACHER QUALIFICATIONS, *TEACHER EVALUATION, NATIONAL INTELLIGENCE NORM, *NATIONAL COMPETENCY TESTS, NATIONAL SURVEYS, *STANDARDIZED TESTS,

THE FEASIBILITY OF DEVELOPING TRADE-COMPETENCY EXAMINATIONS ON A NATIONAL BASIS WAS INVESTIGATED, AND THE POTENTIAL UTILITY OF CURRENTLY DEVELOPED INSTRUMENTS WAS ASSESSED. TWO 1-DAY SEMINARS WERE HELD, AND, AT THE FIRST SEMINAR, FOUR INFORMAL PRESENTATIONS WERE MADE FOLLOWED BY SMALL GROUP DISCUSSIONS. AT THE SECOND SEMINAR, FOUR PAPERS WERE READ AND REACTIONS TO THE PAPERS WERE PRESENTED FOR DISCUSSION. THE COMPLETE TEXTS OF THE PROCEEDINGS OF THE SECOND SEMINAR APPEAR IN THE FINAL REPORT. THE OUTCOME OF THE TWO SEMINARS WAS GENERAL AGREEMENT THAT THE DEVELOPMENT OF OCCUPATIONAL-COMPETENCY EXAMINATIONS ON A NATIONWIDE BASIS WOULD BE A MORE EFFICIENT USE OF PERSONNEL AND SHOULD PROVIDE HIGHER QUALITY EXAMINATIONS. IT WAS ALSO AGREED THAT A PROPOSAL TO DEVELOP TRADE-COMPETENCY EXAMINATIONS ON A NATIONAL BASIS BE PREPARED AND FUNDS SOUGHT TO CARRY OUT THE PROJECT. ALL FEDERAL FUNDS FOR THIS CONTRACT WERE SUBCONTRACTED THROUGH THE BUREAU OF OCCUPATIONAL EDUCATION, NEW YORK STATE EDUCATION DEPARTMENT. (GD)

No. 4

ED012794

FEASIBILITY OF PROVIDING TRADE COMPETENCY EXAMINATIONS FOR TEACHERS ON A NATIONAL BASIS

FINAL REPORT

Project No. 5-0043

Contract No. OE-5-85-110

New York State Education Department

Contract No. C0093

THE DEPARTMENT OF
VOCATIONAL - TECHNICAL EDUCATION
GRADUATE SCHOOL OF EDUCATION
RUTGERS - THE STATE UNIVERSITY

AA000152

FEASIBILITY OF PROVIDING TRADE COMPETENCY EXAMINATIONS FOR TEACHERS ON A NATIONAL BASIS

Project No. 5-0043
Contract No. OE-5-85-110
New York State Education Department
Contract No. C0093

JERALD A. GRIESS
Rutgers - The State University
New Brunswick, New Jersey

FEBRUARY 1967

The research reported herein was performed pursuant to a contract with the Office of Education, U. S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U. S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

ACKNOWLEDGMENTS

Thanks are due to Drs. Donn Billings, Edward K. Hankin, Paul V. W. Lofgren and Mr. Ray A. LaBounty for their informal presentations made at the September seminar.

I would also like to thank Drs. Joseph T. Impellitteri and Benjamin Shimberg, Messrs. Ray A. LaBounty and Joe L. Reed for the papers they presented at the December seminar; and Drs. Edward K. Hankin, Donn Billings, Paul V. W. Lofgren and Ben S. Vineyard for presenting their reactions to these papers. The full texts of these papers and reactions are included as an Appendix to this Report.

INTRODUCTION

The administration of education faces the problem of finding high quality teachers in ever greater numbers. For vocational education the problem is particularly acute. A major contributor to the shortage of vocational teachers is found in the requirements of certification of these teachers. Certification requirements for teachers in trade and industrial areas specify that an individual must have a certain number of years of experience in the occupational area he will teach. Frequently, by the time an individual has acquired the required experience, an occupational change to teaching would result in a lowered salary—a sacrifice many otherwise qualified personnel are not willing or able to make. Furthermore, such a shift would require retraining of the individual (Kazanas and Kieft 1967).

More and more vocational education programs are offered in the public school setting. Federal legislation has stipulated that in order for schools to receive Federal reimbursement, instructors of vocational subjects must be experienced craftsmen. At one time one could assume that the experienced craftsman was competent in all facets of his trade. Today's occupational specialization has tended to alter this picture. Men who have had a well rounded repertoire of skills upon completion of formal training have seen many of their unused skills nearly disappear as they developed other skills to a finer edge of perfection. This poses a problem for vocational education. Vocational courses must be taught by experienced craftsmen, but it is considered desirable for a vocational teacher to be proficient in all aspects of his occupation (Shimberg 1966). The development and use of occupational competency examinations might be one method of contributing to a solution of a number of problems in these areas. A satisfactory score on one of these examinations may be used to short-cut the number of years of experience now required of vocational teachers; it may also be used to help verify a teacher's competency in all facets of his trade.

The Question

The present project concerned itself with two questions: (1) To what extent are trade competency examinations being used in the various states? and (2) Which states would be interested in using trade competency tests that have been professionally prepared and made available through a national clearing center?

Trade competency examinations, in this context, refer to those examinations used to measure a teacher's knowledge and competency in a given trade or occupation. These examinations usually consist of a written phase and a performance phase. Each examination is developed for a specific trade or occupation. Data from examinations are used for such purposes as selection and certification of teachers and for granting of college credit by examination.

A number of states—notably New York, Pennsylvania, Florida and California—have introduced competency examinations under the certification process. A survey in 1959 by Schaefer showed that 16 states were then using tests to evaluate trade competency. However, recent discussions, with officials responsible for vocational education and teacher certification in several of these states, revealed that considerable dissatisfaction existed with the quality of the tests available. By and large these tests are of the paper and pencil variety, although some require actual performance in simulated job situations. There is little evidence to indicate that test questions were pretested or that the instruments were validated according to acceptable test development procedure. Little attention seems to have been given to important technical considerations such as reliability of scores or the objectivity of scoring procedure.

Objectives

The objectives of this project were: (1) To investigate trade competency examination programs now in existence throughout the country; (2) To identify the foreseeable problems of developing trade competency examinations on a nationwide basis; (3) To construct guidelines for the development of trade competency examinations for use on a nationwide basis; (4) To investigate the extent to which states would be interested in using trade competency tests that have been professionally prepared.

Methods

State directors of vocational education or the professional equivalent in each state, District of Columbia, Virgin Islands, and Puerto Rico were contacted concerning the project and were each invited to recommend an individual from his state to serve as a delegate to the two one-day seminars. Four consultants, four reactors and fifteen participants were then selected by the project staff: The main ends of the project were carried out in two one-day seminars. At the first seminar, held on September 19, 1966, it was hoped that some of the problems which might be encountered in the examination development would be identified. We also planned to discuss the practicability of developing nationwide competency examinations and to determine the types of information needed to understand the present examination programs in various states. Four informal presentations were made at the seminar. These presentations highlighted the work on occupational competency examinations currently being done in the speakers' home states. A considerable portion of the time was spent in small discussion-groups with a summarization of each group's deliberations given before the entire group and followed by general discussion. It was anticipated that the comments and recommendations obtained from these discussion groups would provide a basis for the further development of trade competency tests. At the end of the first seminar, assignments of papers were made for the follow-up seminar in December. The latter seminar was held on December 16, 1966. Papers given in this instance attempted to deal with solving the problems which had been identified at the earlier seminar. It was hoped that the papers would include comments on innovations, methods of avoiding current pitfalls, and suggestions, relative to the assigned topic, for developing trade competency examinations for teachers on a nationwide basis. A reactor was asked to respond to each of the papers. Following the presentations there was a general discussion which focused on ways of gaining financial support for examination development. It was agreed that such development was desirable and that occupational competency examinations would be of great value to vocational education.

An instrument to collect data about current competency examination programs was not developed due to two publications which became available after approval of this project. These publications were: Kazanas, H. C. and Kieft, L. D., "An Experimental Project to Determine More Effective Vocational Teacher Certification Procedures in Michigan by Competency Examinations," Department of Industrial Education, Eastern Michigan University, Ypsilanti, 1966. Lauda, Donald Paul, "Factors Related to the Granting of College-University Credit for Trade and Industrial Experience in Institutions Offering Industrial Education," Department of Education, Iowa State University of Science and Technology, Ames, 1966.

Results

The outcome of the two one-day seminars clearly indicated that there was general agreement that the development of occupational competency examinations on a nationwide basis would be a more efficient use of personnel and should provide higher quality examinations. Almost unanimous agreement, that these examinations would be used, resulted. Some states indicated that they would prefer to use data from these examinations for granting college credit and some states would prefer using them for the verification of a teacher's competency. A number of seminar participants expressed the hope that certification requirements would be changed as a result of nationwide examinations so that fewer years of experience would be required.

Discussion

A variety of uses for competency examinations was suggested in the September seminar. Some states would still require a minimum number of years of experience and use an examination only to verify a teacher's knowledge of his occupational area. Other states would use the same test in lieu of the years-of-experience requirement. Still other states would use competency examinations for granting college credit. Such credit, hopefully and presumably, would encourage or assist the individual to complete a degree program. Instances are on record where college credit is, in fact, currently being granted from some trade experience. Another reason for pursuing the examination construction is that it would hopefully reduce the cost to the individual sitting such a test, since developmental costs would be shared across a wider base than would be the case if each state pursued its own aims and objectives. It is essential to stress that, even should occupational examinations be developed on a nationwide basis, each state would determine how and when to use them, if at all.

There are several advantages to be gained by developing occupational competency examinations on a nationwide basis. Many states lack the personnel and the financial resources to develop such examinations individuallly. Furthermore, by pooling resources duplication of effort and replication of errors would be eliminated at least in part. Such an effort would not only be more efficient, but examinations of higher quality should result, since experts from across the nation would be available, providing a broader range of experience in test construction. Occupational examinations developed on a national basis would also be standardized on a national basis, although again it must be stressed that each state may develop its own norm should it so desire. Since few will deny that our society is becoming increasingly mobile, such standardization procedures, by hopefully simplifying certification reciprocity between states, could be of immense appeal to any teacher faced with the prospect of moving to a different state thus necessitating recertification.

Three types of examinations were discussed at the September seminar: written, oral and performance examinations. The written section would undoubtedly be a multiple-choice examination designed to test the candidate's knowledge of his occupation. An oral section was recommended as a means of testing the candidate's ability to communicate his knowledge. It is believed that the possession of such an ability is necessary if one is to be an effective teacher. Since it is believed that a vocational teacher needs performance skills in, as well as knowledge of, his occupation it was decided that the inclusion of a performance section of an occupational examination would be imperative.

At the December seminar, it was suggested that delegates and guests approach their States Departments of Education, college and university Deans and Vocational Teacher Educators urging them to write to Dr. Griess indicating their interest in, and support for, further projects to develop occupational competency examinations. At this point letters have been received from several states indicating a need in their states for these examinations, hope for development of such examinations, and willingness to help in the development. In one case it was suggested that: "Perhaps some of the states may be willing to contribute funds to underwrite such an undertaking. If this be the case, you can count on . . . to assist with such funds which may be reasonable and available." If this is the feeling across the nation, this project should, then, become a reality.

Conclusions

A. It was recognized that there might be difficulties in establishing standardized conditions for administering performance examinations and obtaining reliable ratings.

B. The cost of developing examinations is likely to be substantial especially if alternate forms are needed to preserve security.

C. Evaluating performance sections also poses many problems and is likely to be expensive.

D. There may be a need for specialized norms for various regions or for various specialized groups. The problem of regional differences should be investigated early to ascertain whether or not there is likely to be a serious problem in nationwide examinations. Some matters to consider are: (1) Common core content among geographical regions having specific units for selection by region; (2) Geographical regional differences in the competencies of the skilled craftsmen; (3) Vocabulary differences by geographic area; (4) The fact that standardization at the national level may not meet local needs.

E. The location of examination centers may pose serious problems. Many centers would give rise to administration difficulties and standardized conditions. However, centralized centers, one or two per state, might pose serious difficulties because of the distance candidates would have to travel.

F. It is possible that requiring the candidate to pay a substantial fee would give rise to serious policy questions and might also deter good potential teachers from taking the examinations. The construction of the examinations on a national basis should, however, reduce this cost to a minimum.

G. The question was raised about how much cognizance the national program should take of curriculum changes (such as work with cluster concepts). It was felt that as new curricula gained acceptance special examinations could be developed as was done for new physics programs. Examining on commonalities or clusters within occupational groupings may tend to pull vocational teachers and programs together. This may also assist in developing common names for course offerings and curricula.

H. Continual revision would be necessary to keep examinations abreast of technological change.

I. Difficulty may arise in arriving at a single score to evaluate the competency of an individual, especially when an occupation covers a broad range of knowledge and skills.

J. The competency level of teachers may differ between high school and post-secondary instructors.

Two interesting questions which might be answered by the subsequent evaluation of the use of these examinations are: "Will the beginning worker who has obtained a certificate of completion from a vocational-technical program perform better on the competency examination than a worker who has spent an equal amount of time on the job and not in a formal training capacity?" and "Which aspects of occupational competency are related to experience and training factors and which are not?" (Impellitterri 1966). The value of these examinations can be determined only after using them for a number of years when comparisons can be made between teachers who are certified by examination, eliminating or shortening the years-of-experience requirement, with teachers who are certified by current methods; and by comparing teachers who have been granted college credit by examination with those who have actually taken an equivalent number of credits.

Implications or Recommendations

In order to develop the specific knowledge and skills essential to an occupation, committees would perform analyses of the occupations for which examinations are to be developed. There would be a committee for each occupation. Committees would include employers or supervising personnel directly involved in the occupation, labor union representatives (where applicable), state licensing board members (where applicable), and vocational teachers of the content area. Qualified consultants would augment the committees. Upon completion of the job analysis qualified individuals from the occupation who are not directly involved with the preparation of the analysis, would be asked to review it and assign relative weights to the subject matter included. The individuals conducting the review would represent differing geographic areas. At this stage of the examination development, professional test specialists will meet with the analysis committee to develop the specifications for construction of an examination. The specifications will be carefully reviewed before actual examination construction begins. Recognizing both the importance of performance examinations and the difficulties posed by existing techniques, the delegates considered the following possibilities as worthy of exploration:

A. Use of semi-finished or partially completed tasks to conserve time and focus attention on the critical skills.

B. Simulation—use of trainer type devices, electronic models, etc., which simulate conditions without requiring performance on live work.

C. Sampling—such as requiring a candidate to cut only a few teeth in a gear not the whole gear.

D. Use of stop action on film or video tape to show applicant a critical operation and require him to tell what to do next, what will happen, what is wrong, etc.

SUMMARY

The purpose of this project was to investigate the feasibility of developing trade competency examinations on a national basis and to assess the potential utility of developed instruments.

Two one-day seminars were held with delegates from 23 states participating. At the first seminar, four informal presentations were made and were followed by small group discussions. At the second seminar, four papers were read and were reacted to and their implications discussed.

The outcome of the two seminars indicated that the development of occupational competency examinations on a nationwide basis would be a more efficient use of personnel and should provide higher quality examinations. Almost unanimous agreement that the examinations would be used, resulted. Some states indicated that they would prefer to use data from these examinations for granting college credit and some states would prefer using them for the verification of a teacher's competency. A number of seminar participants expressed the hope that certification requirements would be changed as a result of nationwide examinations so that fewer years of trade experience would be required, thereby increasing the pool of qualified personnel. One result of this increase would be to make available more prospective teachers in a time of acute shortage.

It was the consensus of the group that a proposal to develop trade competency examinations on a national basis be prepared and funds sought to carry out the project. It was suggested that delegates urge their state and university administrators and staff to submit letters indicating willingness to cooperate in such a project. A number of supportive letters have been received expressing need for and willingness to use these examinations, and a desire to facilitate their development.

BIBLIOGRAPHY

Impellitteri, Joseph T., *Constructing Valid Occupational Competency Examinations*. Unpublished paper. December, 1966.

Kazanas, H. C. and Kieft, L. D. *The Case For Certification By Competency Examination*. School Shop, 1967, 26-29.

Shimberg, Benjamin. *The Performance Phase of Trade Competency Examinations*. Unpublished paper, December, 1966.

APPENDIX A

SEMINAR PARTICIPANTS

Frank Adelman	State Department of Education, Arkansas	Vocational-Industrial Teacher Educator
Edward M. Alderman	State Department of Education, North Carolina	Assistant State Supervisor
Lawrence E. Allwardt	State Department of Education, Wisconsin	Supervisor, Industrial Education
Hosmer P. Allyn	Old Dominion College, Virginia	Assistant Professor of Industrial Education
William F. Banaghan	State Department of Education, Iowa	Area School Consultant
George Barton	Educational Testing Service, New Jersey	Program Director for Junior Colleges
Walter A. Bialobrzeski	State Department of Education, Connecticut	Head Teacher Trainer
Don Billings	State Department of Education, New York	Coordinator of Industrial Teacher Education
Clement Bogaard	State Department of Education, Illinois	Trade and Industrial Supervisor
George L. Brandon	The Pennsylvania State University	Professor and Chairman of Vocational Education
Kinneth Chambliss	University of Maryland	Associate Professor of Education, Department of Industrial Education
Vincent Cieri	U.S. Army Signal School, Ft. Monmouth, New Jersey	Chief of Evaluation Division
Theodore J. Cote	Temple University, Pennsylvania	Chairman, Trade and Industrial Education
Richard Elkins	University of Maryland	Assistant Professor of Education, Department of Industrial Education
Joseph Frank	U.S. Army Signal School, Ft. Monmouth, New Jersey	Chief of Instructional Methods
Angelo C. Gillie	Rutgers - The State University	Association Professor of Education, Department of Vocational-Technical Education
Geoffrey D. Gould	Rutgers - The State University	Assistant to the Director of Admissions
Jerald A. Griess	Rutgers - The State University	Principal Investigator, Lecturer in Education, Department of Vocational-Technical Education
Jon Grove	State Department of Education, Kansas	State Supervisor, Trade and Industrial Education
Howard R. Hammond	State Department of Education, Iowa	Consultant, Vocational Teacher Education
Edward K. Hankin	Florida State University	Professor of Education and Chairman, Department of Vocational and Industrial Education
Peter T. Harkness	State Department of Education, New York	Development Progrem Coordinator, Bureau of Occupational Research
James L. Heller	State Department of Education, Delaware	Supervisor, Trade and Industrial Education
Addison Hobbs	Washington, D.C. Department of Education	Supervising Director, Trade and Industrial Education
Joseph T. Impellitteri	The Pennsylvania State University	Assistant Professor of Industrial Education
Mary B. Kievit	Rutgers - The State University	Associate Professor of Education, Department of Vocational-Technical Education
Jack Kleinman	State Department of Education, New Jersey	Supervisor, Apprenticeship Training
Louis Koenigsberg	State Department of Education, New York	
Ray A. LaBounty	Eastern Michigan University	Professor and Head, Department of Industrial Education
Gordon Law	State University College, Oswego, New York	Associate Professor of Education, Division of Vocational-Technical Education
Ralph LoCascio	State Department of Education, New Jersey	Director of Research and Development Branch
Paul V. W. Lofgren	State Department of Education, California	Supervisor, Occupational Proficiency Testing

SEMINAR PARTICIPANTS

Joseph F. Luetkemeyer	University of Maryland	Associate Professor of Education, Department of Industrial Education
Marion E. Maddox	University of Arkansas	Professor of Industrial Education
Harold L. Mailman	State Department of Education, Maine	Director of Trade and Industrial Education Department
L. C. McDowell	University of Kentucky	Assistant Professor and Head,
Withro McEnge	United States Office of Education	Specialist, State Research Division of Adult and Vocational Research Planning
S. Charles Meislin	State Department of Education, New York	Research Associate, Bureau of Occupational Research
Robert Mertens	University of Maryland	Assistant Professor of Education Department of Industrial Education
John B. Moullette	Rutgers - The State University	Lecturer in Education Department of Vocational-Technical Education
Charles W. Nichols	Kent State University, Ohio	Associate Professor of Education, Division of Vocational Education
John L. O'Brian	Rutgers - The State University	Associate Professor of Education, Department of Vocational-Technical Education
Frederick Okula	State Department of Education, Connecticut	
Wilmot F. Oliver	State Department of Education, New Jersey	Director of Vocational Teacher Education
Fred Porges	Middlesex County Vocational and Technical High School, New Jersey	Coordinator of Apprenticeship Training
Joe L. Reed	The University of Tennessee	Professor and Head, Industrial Education
Edward M. Roden	State Department of Education, New York	Syracuse Coordinator, Vocational Teacher Education
Ralph A. Rush	Rutgers - The State University	Lecturer in Education, Department of Vocational-Technical Education
Carl J. Schaefer	Rutgers - The State University	Professor of Education and Chairman, Department of Vocational-Technical Education
C. Paul Sherck	State Department of Education, Missouri	Director of Instruction, Vocational-Technical Education
Benjamin Shimberg	Educational Testing Service, New Jersey	Director of Vocational-Technical Education Projects
Harold Starr	State Department of Education, New Jersey	Director of Progress Evaluation, Vocational Division
Thomas C. Stone	Southern State College, South Dakota	Assistant Professor and Chairman, Division of Vocational and Industrial Education
Peter A. Taylor	Rutgers - The State University	Associate Professor of Education and Statistics
Bruce W. Tuckman	Rutgers - The State University	Associate Professor of Education, Department of Vocational-Technical Education
Walter E. Ulrich	State Department of Education, Utah	Specialist, Trade and Industrial Education
Ben S. Vineyard	Kansas State College of Pittsburg	Associate Professor and Chairman, Trade and Technical Education
James Wilson	State Department of Education, Colorado	Assistant Director of Vocational Education
John C. Wilson	State Department of Education, Delaware	Supervisor, Occupational Information and Career Services

APPENDIX B

ACKNOWLEDGMENT

Special thanks are due the Bureau of Occupational Education Research, the State Education Department, the University of the State of New York for taking an interest in this project and giving it support by making a portion of their Office of Education, U.S. Department of Health, Education, and Welfare contract funds available.

APPENDIX C

TABLE OF CONTENTS

Page

Constructing Valid Occupational Competency Examinations
Joseph T. Impellitteri 1
Reactor - Edward K. Hankin. 24

A Limited Field Test of the Automotive Competency Examination
Ray A. LaBounty 31
Reactor - Donn Billings. 39

Preparation, Administration and Implementation of Trade Competency Examinations for College-University Credit
Joe L. Reed 40
Reactor - Ben S. Vineyard 53

The Performance Phase of Trade Competency Examinations
Benjamin Shimberg 56
Reactor - Paul V. W. Lofgren 68

APPENDIX D

CONSTRUCTING VALID OCCUPATIONAL COMPETENCY EXAMINATIONS

by

Joseph T. Impellitteri[1]

This paper focuses on three questions related to the current effort to examine the feasibility of establishing a nationwide occupational competency examination program.

1. What consideration should be given to reliability and validity in constructing nationwide occupational competency examinations?

2. How may valid and reliable occupational competency examinations be constructed?

3. How may the validity and reliability of an occupational competency examination be measured?

In the discussion which follows no mention of cost nor practicality has been made. Primary emphasis has been placed upon the steps which should be taken and the factors which should be considered in conducting an effective occupational competency examination program.

Validity and Reliability - Their Meaning, Utility and Factors Related to Them

In constructing any test there are two questions one should ask himself. First, will this test be suitable for this specific purpose? Second, will the test scores obtained by the people I test be accurate?

The first question describes validity - the second reliability. Validity tells us if a specific test is suitable for a particular purpose. Reliability is related to the accuracy of the test scores.

In building an occupational competency examination in printing we certainly should be interested in whether or not the test we're constructing is suitable to measure occupational competency in printing. We should not, on the other hand, be interested in whether it's suitable for measuring anxiety. A test is valid only for a specific purpose. A test that is suitable, and thus valid in measuring intelligence would not be valid for the purpose of measuring extent of outdoor activity. The concept of reliability, though, is not related to suitability. A test that is highly reliable is highly reliable, period. A test either yields accurate scores or it does not. Whether the score accurately represents occupational competency in printing or intelligence, or anxiety is not pertinent to reliability.

The Stanford-Binet intelligence test is highly reliable as well as being highly valid for the purpose of measuring intelligence. It would not be highly valid for the purpose of measuring occupational competency in printing.

---

[1] Dr. Impellitteri is Assistant Professor of Vocational Education at The Pennsylvania State University.

The Relationship Between Validity and Reliability.

A test that is valid must also be reliable. Micheels and Karnes have stated:

> It can be immediately seen that reliability is closely connected with the validity of a test. If a test is valid, it must be reliable. That is, if a test measures effectively what it is supposed to measure, then presumably it does this accurately and consistently. At the same time it must be remembered that a test might be highly reliable and still not be valid. (12)

Why is this true? It is based upon the concept of representative sampling. Turn to Table 1 on the following page describing the universe of items which might be included in an occupational competency examination in Electronics. Viewed in this way I think that we can all agree that the number of items which can be constructed in this framework is infinite. If one adequately samples from the universe, taking one representative item from each of the 48 cells in Table 1 then the resulting test should be valid. If we construct two items for each cell the test will be more representative of the universe, and hence more valid. The more items constructed within the framework presented the more valid will the test be.

A test constructed in this manner must also have high reliability. It follows naturally from the procedure. Why is this so?

A reliability coefficient indicates the extent to which the scores obtained by individuals taking the test are representative of their "true" scores. These "true" scores I'm talking about are the scores these individuals would obtain if we were to give them a test including all the items in the universe. But we all know this is an impossible task. We must deal only with sampling of items from the universe. How well these individuals' obtained scores represent their "true" scores is dependent upon the degree to which the selected items in this test represent the universe of items.

Thus, when we talk of the bases, in measurement terms, for validity and reliability of a test, we're talking about the same thing - the extent to which the items on a test represent the universe of items.

What I'm trying to stress at this point in the discussion is the construction of tests with high validity, for if you have high validity you have everything. Conversely, if you have high reliability, you might just have nothing.

Can we safely say then that if we draw up a satisfactory table of specifications and adequately represent the content areas and specific objections included in the table with test items that we'll have a valid and reliable test? This does not necessarily follow. Although these steps are essential to validity and reliability of a test they are not enough. We've all seen beautifully done blueprints for homes with carefully compiled tables of specifications for the carpenters, plumbers, and electrician. But the high quality of the blueprints does not guarantee you'll get a well-constructed house. There still is an essential step missing. That is the implementation of the blueprint, the workmanship involved.

Table 1

The Universe of Items in the Field of Electronics

Objectives*	D.C. Electricity	A.C. Electricity	Tubes and Semi-Conductor Devices	Basic Electronic Circuits	Industrial Electronics
Knowledges					
Comprehension					
Application (Problem Solving)					
Analysis					
Synthesis					
Evaluation					

*From Bloom, Benjamin S. (Ed). _Taxonomy of Educational Objectives_. N.Y.: McKay, 1956.

And so it is with test construction. The item writing must be well done. Ambiguous or confusing items lower both the reliability and validity of a test. At this point I should stress that no amount of statistical manipulation can introduce into the test anything that has not been written into the items; and any validity written into the items will forever plague the efforts of the investigator to analyze the source of discrepancies.

Of what utility then is the concept of reliability? For our purposes I can think of two ways in which estimation of reliability would be beneficial. First, if the measured reliability were low one could be assured that the validity of the test was low. Secondly, if one were reasonably sure that the test had high content validity the reliability coefficient could then be used to interpret the scores in terms of the confidence one can have in the test results (See Appendix A).

The Types of Validity.

There are essentially three types of validity: content validity, construct validity, and criterion-related validity. I think we need focus on the first two only. Criterion related validity is appropriate for tests which are designed to be used to forecast consequent behavior, as exemplified by aptitude tests. In these occupational competency tests we wish only to identify what knowledge and skills has an individual acquired in this occupational area.

Because I have taken a stand at this point on eliminating criterion related validity from the scope of this discussion I feel I must now justify it. Many of you might be saying to yourselves at this point, "I wish we could use the results of this test to accurately predict the extent of an individual's teaching effectiveness. I think predictive validity is important to consider." I must make a plea at this point to confine ourselves strictly to a discussion of measuring occupational competency. That particular job is quite extensive and complex enough without considering a broader focus. Ask Mr. Lofgren particularly, and several others in this room if they think we have a big enough job to do. Predicting teaching performance is most certainly a highly significant problem in considering the entire task that must be done. I do submit, however, that we've taken a giant stride already in undertaking only the occupational competency measurement. The exclusion of criterion-related validity from the discussion does not appear to be disastrous at this time.

What about construct and content validity? What implications do these two concepts have in considering occupational competency testing?

In my discussion of the universe of items and the necessity of an adequate sampling of these items to insure validity, I was referring to content validity. That is, given a table of specifications which describe some framework of pertinent behaviors, content validity is concerned with the adequate sampling of these behaviors in a test designed to measure these behaviors. Content validity, in other words, tells us something about the adequacy of the test as representing a domain of behaviors such as occupational competency in electronics.

Construct validity, on the other hand, tells us little about the validity of the test itself. The focus of construct validity is upon the validity of the table of specifications itself. Is the domain of behaviors I have outlined psychologically meaningful? That is, if I wish to measure occupational competency in electronics have I adequately defined the behaviors which would be exhibited by a highly competent electronics expert in my table of specifications? Construct validity is focused on the process whereby the pertinent trait or characteristic such as occupational competency in electronics is defined in terms of specific behavioral objectives.

Validating the Written vs. Performance Tests.

In discussing the validation of occupational competency examinations it must be decided whether to look at the written and manipulative parts separately or to consider them together.

In establishing the content validity of both parts of the exam the decision is irrelevant. That is, content validity is neither improved nor lessened by separating the two parts as opposed to considering them together as a whole. A glance at Table 2 will reveal why this is so.

The focus of content validity is on representative sampling of the universe of items. Thus, whether we work with "knowledge of terms" and "understanding of physical principles" together with or separate from "ability to work within specified tolerances" is irrelevant. The representativeness of the selected items should be the same. Since unique measurement problems enter into the assessment of manipulative-performance tasks it probably would be most beneficial to consider the two parts separately. One example of these unique problems is degree of sampling. With a well-constructed paper and pencil test it is possible to measure 200 to 300 relatively independent items of behavior in a three to four-hour testing period. During the same time period, however, only 10 to 20 manipulative-performance behaviors may be observed and measured. Cronbach has stated that:

> Low reliability is characteristic of worksamples where one error may disturb the entire sequence of performance, and several samples of performance must therefore be obtained. The more successful . . . tests usually include a large number of short, similar items, rather than a few complex sequences of performances (4).

Evidence of Construct Validity.

What should be the magnitude of a correlation between a paper and pencil test and a manipulative-performance test in the same occupation? Let us first examine the extremes. If the correlation approached the limit of 1.00 there would be no necessity for using both tests. They would both be measuring the same thing.

Suppose, on the other hand, the correlation was found to approach .00? Is this the ideal situation? One certainly could say, on the basis of this finding, that the two parts of the test were measuring different aspects of competency in the trade. I would, however, question such a finding. I would suspect the written test, the performance test, or both parts as possessing low content validity.

My rationale for such a suspicion would be that one must possess some knowledge and understanding of principles involved in the occupation in order to be able to adequately perform tasks representative of that occupation. A plumber need not, perhaps, know the temperature at which solder melts, but he should know that heat must be applied to a certain area of a copper fitting in order that the solder applied will be drawn and make a tight joint with the copper pipe.

Somewhat arbitrarily I would choose as an acceptable correlation some magnitude in the range of .30 to .60.

Table 2

A Partial Simplified Table of Specifications

Machine Shop

Measurement Objectives	Content Areas		
	Engine Lathe	Drilling & Boring	Shaper
Knowledge of terms			
Understanding Physical Principles			
Ability to work within specified tolerances			

The Effect of Guessing on Reliability and Validity.

An interesting alternative to overcome errors in scores introduced by examinee guessing has been introduced by Rembert R. Stokes in the January, 1966, issue of the Phi Delta Kappan. What he suggests is to introduce the "split-response technique". To give you an example of how this technique works let's look at a typical multiple-choice test item.

$$\sqrt{16^2 \times 12^3} \;/\; \sqrt{2}$$

a. $16\sqrt{12}$

b. 16

c. 192

d. $48 / \sqrt{12}$

e. none of the above.

From the point of view of the examinee, what process does he go through in answering the question? He might be able to solve the equation immediately and circle alternative "e". If he cannot solve the equation entirely he might at least be able to reduce the number of possible alternatives. For instance, he might be able to estimate the answer to be 100+ without being able to go through the solution. He then eliminates alternative "a" and "b" from consideration. He might guess from the three alternative "d" and be penalized for marking the wrong answer. How does this differ from the person who knows nothing about square roots and consequently cannot even estimate the answer but guesses "d" at random? Traditionally, we have had no way of distinguishing between these two responses.

Mr. Stokes has suggested that the following scheme could be used. First, indicate to all examinees that each item is worth ten points. Then, allow them to assign points from the kitty of ten to each of what he considers to be possible alternatives. Thus, the person who was sure that alternative "e" was correct in the above item would place the ten points on alternative "e". The second individual would probably divide up his points equally between alternative "c", "d", and "e". In the third case the examinee having no basis to act differently would place 2 points on each of the five alternatives.

This technique is the only one I've come across to adequately account for differences in scoring between individuals who know nothing about an item and those who have sufficient understanding to eliminate one, two or three of the five alternatives. The first person would have received the highest score possible on the item - ten points. The third individual would be credited with two points - that which we'd expect by chance. The second individual would receive three to four points - more than could be expected by chance alone. I'm convinced that this kind of an approach would increase the validity and subsequently the reliability of these competency examinations. Having reviewed several measurement text authors' comments on correction for guessing on examinations (1, 2, 4, 12, 14) it is apparent that little agreement now exists. No concensus exists as to whether or not to use corrections for guessing, and whether or not to tell the examinees to avoid guessing.

The Validation of Part-Scores.

Because of the numerous complex skills and knowledges to be tested within any one occupation there should exist several relatively independent meaningful aspects of competency in the occupation. These aspects of competency should in addition, be measurable.

By utilizing the table of specifications as a logical framework for the clustering of items, some meaningful divisions within the test should emerge.

I assume that we all agree that occupational competency in a specific work field is not a unitary ability. It is a complex organization of a number of unitary abilities. It is handy for us to use the term "occupational competency" as if it were a unitary ability. We speak of some persons as being more or less competent carpenters than others. Individuals, however, do not possess a degree of competency in carpentry. They do possess certain manipulative skills, knowledges and the ability to coordinate these when applied to certain tasks. A global score, then as used to describe what we conveniently call occupational competency is somewhat misleading.

Part scores would appear to be most useful in terms of evaluation of an examinee's performance as well as its diagnosis. For instance, a person could do quite well on a well-constructed occupational competency examination in plumbing yet know nothing about blueprint reading and layout of a job. I contend that it is important to know the various strengths and weaknesses of an examinee's performance, not merely his total score.

If part scores are utilized, however, much effort should go into their validation - not only in terms of content validity as was discussed previously. Some evidence of construct validity must also be collected. Data regarding the intercorrelation between the scores should be collected. If Part I correlates with Part II .95 and with Part III. 89, the part scores on this test would be useless. The goal should be to construct parts within a test so that correlations between them are no higher than .20 to .30. This kind of evidence would indicate that the parts of the test are actually measuring different aspects of competency.

The Establishment of Norms

In itself a raw score obtained on a test is essentially meaningless. If an individual obtains 140 right out of a total of 200 items on a written carpentry test, what can we say of this individual's competency in carpentry? We can interpret the raw score as a percentage of the total items correctly answered - in this case, 70 per cent. Is this percentage passing or failing, good or bad? In fact, is a score of 50 per cent bad, or might it be good?

I contend that there is no way of discriminating the passing score from the failing score, the good score from the bad score, no matter what the individual score may be, except through the establishment of norm data. No individual is a completely competent carpenter or a totally incompetent carpenter. A person's degree of competence in an occupation should be based upon relative positioning.

The establishment of norms will allow for the meaningful interpretation of exam scores. Cutoff scores for passing or failing could be assigned in terms of percentiles instead of percentage of total items correctly answered. One could arbitrarily set the 50th percentile as the cutoff for passing. That is, an individual must be of at least average competency in his occupation to pass. Another person with a different orientation and background could establish a cutoff for passing at the 75th percentile. In the latter case an examinee must fall within the top one-quarter of workers in a specific occupation in order to pass. The standard could vary widely between states and between institutions, but at least there would be some uniformity in the meaning of the obtained scores when reported as percentiles.*

The Norm Group.

In establishing norms primary consideration must be given to the manner of selection of the individuals who are to be included in the norm group. A decision must be made as to the nature of the persons in this norm group. The basis for this decision lies in the answer to the question, "With whom do we want our prospective candidates to be compared?" This answer is not an easy one. Consideration must be given to a variety of factors.

Geographical Factors.

Do we desire to have scores attained by prospective examinees in carpentry living in Altoona, Pennsylvania, to be compared with scores attained on the same examination by carpenters in the same general locality, by carpenters across the state, the region, or the nation? Is there enough variability in this occupation to eliminate nationwide comparison or even, perhaps, statewide comparison? What about other occupations such as mechanical drafting, computer programming, or chemical technician? Does the variability from state to state diminish, or increase.

I won't even attempt a partial answer to either of these questions or to the hundreds of related questions that might arise. The important implication is that this is at least one factor which must be taken into consideration in building norms. The same factor must be taken into consideration before any one specific competency test is ever built.

Experiential and Training Factors.

In addition to the geographical representation of the norm-group, its level of training and experience must be in some manner decided upon. Again the crucial question is, "With whom should potential examinees in an occupation be compared?" Should their test scores be compared with test scores attained by a representative sampling of journeymen only? What about non-apprenticeable occupations -- those with at least six years of experience in the occupation? Should apprentices also be included in the norm group, or trainees? What kind of representation should there be in a norm group for an occupation?

---

*There is some opposition in the tests and measurement literature to the practice of establishing percentile norms. Some standard score system like T scores could be used. For the sake of this discussion, the author has used the more commonly known percentile norms.

The Construction of Norms for Part-Scores.

In a previous section of this paper on the use and validation of part-scores of a test it was recommended that part-scores be utilized. If, in actuality part-scores are introduced in addition to the global score, constructing norms for the part-scores becomes essential. Why must this be done?

Table 3 is a hypothetical and simplified table of percentile equivalents for a hypothetical occupational competency examination in electronics. The hypothetical examination consists of four parts of 50 items each. If one were to use the part-scores but not establish norms for them much important information would be lost. Glancing at the table for a moment, what information would be lost? Suppose an individual scores at the 50th percentile according to his total score (the last column in the table). Suppose also that the percentile equivalents from the part-scores were not available. Then one looked at the raw part-score attained by the individual. Say they were 30, 30, 30 and 31. The only interpretation one could make would be that this individual was equally competent in each of the four parts. Looking at the percentile equivalents for the part-scores, however, tells a different story. With the given raw scores, one could say that this individual possessed average competence in AC electricity, was extremely competent in the area of DC electricity, was very poor in communications systems and was somewhat above average in tests and measurements. This information is crucial and should be acquired.

Table 3

Percentile Equivalents: Electronics Examination
Reduced and Simplified

Percentile Equivalents	Part Scores (Raw)			Total Raw Score (200 Items)
	A.C. Electricity (50 Items)	D.C. Electricity (50 Items)	Communications Systems (50 Items)	
99	.	.	.	.
95	.	.	.	.
90	41	25	46	.
75	37	21	45	.
50	30	15	41	121
25	23	9	37	.
10	19	5	36	.
5	.	.	.	.
1	.	.	.	.

A Plan for Constructing Quality Occupational Competency Examinations

The plan to be described has been set up in the belief that any test construction program has to be planned as an integral unit. This means that the definition of the domain of skills and knowledges to be tested, the sampling of that domain, the construction of items, the design for administration and scoring, and plans for establishing validity, reliability and item difficulty of the test must all be considered together. No single step or phase can be planned in isolation. Unless these problems are all attacked together from the outset of the project, no scientific measuring device can result.

Consideration was given to the publication Standards for Educational and Psychological Tests and Manuals (15) in devising the proposed plan as presented below (See Appendix B).

Construction of the Test.

When the need for an occupational competence examination in a specific occupation has been established by a representative committee the following steps should be taken:

1. An occupational committee consisting of from five to nine recognized experts in the occupation representing interested geographical regions will be employed to meet with an occupational specialist and a test construction expert for a period of time. The efforts of the committee should be directed toward the determination of those skills, knowledges, understandings, and other abilities which should be possessed by a competent worker in the occupation, and the construction of a suggested plan for evaluating those abilities.

2. When the domain of pertinent occupational behaviors has been decided upon, each regional representative of the committee should make available a copy of the document containing the agreed upon description of the domain to the responsible vocational-technical administrator/s in each of the states he is representing. An additional meeting of the occupational committee will be necessary in order to communicate the extent of agreement or non-agreement. Once the table of specification as described above has been accepted the major job has been accomplished.

3. At least three of the committee members in cooperation with the test construction expert should write items sampling the accepted domain of behaviors. At least three times the number of items that will eventually be used should be written -- probably 500 to 1,000 items. This group should be responsible for constructing representative manipulative performance tasks as well as paper and pencil items.

4. An acceptable scoring key for the written items should be devised. The scoring of the manipulative-performance tasks should be devised in as objective a manner as possible. A scoring scheme similar to

the one proposed by Fleming and Hankin (7) should be constructed. The importance of objectivity in performance testing has been stressed by Fatter and Medley (6). They state that, "Objectivity is necessary because only to the degree that a test is objective can it measure anything that is a trait of the individual being measured. If two scorers score the same individual in two different ways, the test is to a degree measuring the observer instead of the man being tested." Objectivity is gained be breaking down a complex task into specific components of that task. The more the scoring scheme focuses on specific observable behavior, the less likely is subjectivity to be of concern as a source of significant error.

5. Two parallel 300 item tests as well as two performance tests should then be administered to a group of representative workers in the occupation as well as to teachers of the occupation as suggested by Kazanas and Kieft (9). Critical comments of the examinees should be encouraged. The group of selected trial examinees should represent the geographical regions in accordance with the occupation committee's representation. Some logical scheme for deriving part-scores should also be constructed.

6. On the basis of an empirical analysis of the test data (see next section for further description) and critical comments of the examinees the item writing committee as described in step #3 should evaluate the results, and revise the examinations (also described in the next section of this report). The final forms of the written test should include no more than 200 items.

7. The next step would be the collection of norm data. At least 500 to 700 workers in an occupation would compose the standardization sample the final number depends upon: 1) the variability of the occupation from region to region; and 2) the range of regions to be represented. Regional norms and nationwide norms should be constructed for each part and the total score of each form of the exam.

8. A test manual should finally be developed including: Directions for administering and scoring both forms of the written and performance exams; the table of specifications constructed for the examinations; the regional and nationwide norms; a description of the standardization sample; and suggested interpretation of the test scores.

Obviously, strict security of these exams must be maintained, or else the test results would be worthless. There are many acceptable procedures for insuring security of exams, but it is outside of the scope of this paper to undertake a discussion of them.

Recommended Procedures for Measuring the Validity and Reliability of Occupational Competency Examinations

Content Validity.

Much of the discussion in this report has dealt with the necessity of establishing the content validity of occupational competency examinations. The procedure which has been described for building content validity into these examinations has been in constructing items representing a domain of skill and knowledges in an occupation. The objective was to determine the extent to which an individual possessed the knowledges and skills necessary for competent performance in the occupation. This procedure may be described as a logical keying procedure. This approach is exemplified by identifying what an individual needs to know, and what he needs to do in an occupation and then develop an instrument for measuring these knowledges and performances.

Statistical procedures for measuring content validity are non-existent. The process is one of critical examination and judgment.

Construct Validity.

There are several analyses which can be conducted which will provide evidence as to the construct validity of an occupational competency examination. The process of examining construct validity involves systematic investigation of the numerous variables which are related to occupational competency.

Empirical Keying.

In contrast to the logical keying approach utilized in the discussion of content validity, empirical keying provides a different kind of information. The implication of empirical keying for the purposes of competency examining is exemplified in the following procedure.

Given a 200-item written test in carpentry, administer it to a group of journeymen carpenters and as well as to a group of journeymen in other construction occupations. The results could be analyzed in accordance with the structure presented in Table 4. One could, based on these test results be able to identify items similar to #3 and #5 in Table 4. These items, as one can readily see appear to be unique to the carpentry trade. Items similar to #1, #2, and #200 might be included in a separate examination for construction trades.

What could result from this approach is a 150-item test for all construction occupations in addition to 50 items in a specific occupation. The focus in such an effort would be on avoiding duplication of effort from occupation to occupation.

Independence of Part-Scores.

One basic question arises when constructing tests designed to provide part-scores as well as a total score. The question is, "Are the part-scores

Table 4

Item by Item Analysis of Results of
Carpentry Examination

Item #	Percentage of group passing item Carpenters	Other Construction Workers
1	90%	85%
2	60%	65%
3	85%	26%
4	25%	48%
5	76%	30%
:	:	:
200	25%	18%

independent?" If they are not relatively independent as evidenced by low intercorrelations between them, they are practically useless. A simple correlation analysis is the only necessary step to determine this degree of independence.

Another empirical measure of the utility of part-scores would be provided by a factor analysis of the items of the test. That is, the items that seemed to be measuring a common aspect of occupational competency could be identified. If these common factors closely corresponded with the logically constructed part-score configuration, then the part-score framework would be empirically verified.

What Factors Are Related to Scores on a Competency Examination?

It has been suggested that the standardization group utilized in establishing norms for the competency examinations should be composed of a representative group of at least journeymen level workers in the occupations. Several interesting questions with respect to construct validity can be posed at this point. What would happen if the test were administered to apprentice level workers in an occupation - or to vocational or technical students in the occupation? Will a beginning worker who has obtained a certificate of completion from a vocational-technical program perform better on the competency examination than a worker who has spent an equal amount of time on the job, and not a formal training capacity? Which aspects of occupational competency are related to experimential and training factors, and which are not?

The answers to the above questions, and to similar questions which may be posed are crucial in examining the construct validity of occupational competency examinations. A design for the analysis of test data which directly relates to some of the questions posed above has been included as Appendix C to this report.

Reliability.

Since it has been decided that a reliability coefficient would be useful to calculate the recommended method is presented here. If one of the preceding recommendations of the report were to be utilized, that of constructing parallel forms of the examination, both written and performance, then calculation of reliability is both simple and appropriate. The two forms of the examination are administered to the same group. Correlations are then calculated between the corresponding part-scores in the two forms as well as the total scores. The correlations are the reliability coefficients of the test itself and of the part-scores.

How, then would the reliability of the performance exams on the two forms be calculated? A recommended procedure would be to follow Hoyt's* Methodology of determining reliability through an analysis of variance. The three sources of variation would be differences between the mean scores of individuals in the group, differences between the mean of the two forms, and an error component. The formula which could be utilized is:

$$\text{reliability} = \frac{A - B}{A}$$

where: A = mean square, differences between individuals

B = mean square, error

Item Analysis of Test Results

Typical item analyses are designed to provide the test constructor with information as to the contribution of each of the test items to the total test score. The performance of higher achievers on the test and the lower achievers on the test is compared item by item. The upper 27 per cent of the group and the lower 27 per cent of the group are usually selected for comparison.

The comparisons proceed in a manner similar to the structure presented in Table 5. The rationale is that persons who score higher on the test as a whole should score higher on each item than persons who do poorly on the test as a whole. Items 1 and 2 in Table 5 reflect these expected differences, but items 3, 4, 5 and 200 do not. Items such as the latter should be reviewed by the test constructor in order to determine the source of these differences.

An item which is extremely easy or extremely difficult of course will not discriminate, nor is it expected to. Often these items are included for other purposes. However, the results obtained on item #200 are quite suspicious. The item would probably be discarded.

The objective of item analysis is to improve the measuring instrument. It is possible to pick out ambiguities in items, misleading words, and totally irrelevant questions. It is a useful procedure and should most definitely be considered in a large scale testing effort.

---

*Hoyt, C. T. "Test Reliability Estimated by Analysis of Variance," Psychometrika, VI (1941), pp. 153-160.

Table 5

An Item Analysis of Test Results

Item #	Per Cent Passed Upper 27%	Per Cent Passed Lower 27%	Discrimination Index	Difficulty Index
1	80	40	.40	.40
2	90	20	.70	.45
3	36	50	-.14	.57
4	66	60	.06	.47
5	98	98	.00	.02
:	:	:	:	:
200	40	80	-.40	.40

Concluding Comments

This paper has been written for the purpose of presenting in an organized manner the technical considerations which should be respected in attempting to build high quality occupational competency examinations. I am sure that all of the recommendations included in this report will not be incorporated in the final effort, if and when it is initiated. Nor was it meant that they all should be so used.

Hopefully, this report will, however, awaken other, more creative ideas for attacking the kinds of problems I have been discussing today.

APPENDIX A$_{D}$

The Utility of the Reliability Coefficient

I. Reliability as related to validity.

An inaccurate test (low reliability) cannot have high validity. The validity can be no higher than the square root of the reliability coefficient. Example: If $r_{1I} = .49$, the validity can be no higher than .70*.

II. Interpretation of scores.

The standard error of measurement of a test may be derived directly from the reliability of the test and can be used to interpret test scores.

$$\sigma_{\text{meas}} = \sigma_1 \sqrt{1 - r_{1I}}$$

where: σ_1 = standard deviation of scores on test

r_{1I} = reliability coefficient

$$x_{\text{true}} = r_{1I} \; x_{\text{obt.}}$$

where: $x_{\text{obt.}}$ = the deviation of the score of an individual from the group mean

Suppose Joe obtains a score of 110 on the written examination we've administered. Reliability of the test = .90, and $\bar{X} = 100$, and SD = 15. Joe's $x_{\text{true}} = -9$; $100 \neq 9 = 109$; $\sigma_{\text{meas}} = 5$.

Ninety-nine times out of 100 Joe's score will fall between 96 and 122.

---

*Cronbach, page 132, Anastasi, pp. 129-131.

APPENDIX B$_D$

Standards of test construction which should be taken into account in a nationwide occupational competency testing program*.

A. Dissemination of Information.

1. When a test is published for operational use, it should be accompanied by a manual (or other published and readily available information) that makes every reasonable effort to follow the recommendations in this report. ESSENTIAL

2. The test and its manual should be revised at appropriate intervals. While no universal rule can be given, it would appear proper in most circumstances for the publisher to withdraw a test from the market, if the manual is 15 or more years old and no revision can be obtained. (Comment by J.T.I.: Because of the nature of the competencies which are being measured in occupational competency examinations, five to seven years would be a more appropriate interval.)

   2.1 Competent studies of the test following its publication, whether the results are favorable or unfavorable to the test, should be taken into account in revised editions of the manual or its supplementary reports. Pertinent studies by investigators other than the test authors and publishers should be included. VERY DESIRABLE

   2.4 When a test is issued in revised form the new copyright date should be indicated on both the test and the manual. The nature and extent of the revision and the comparability of data between the old test and the revised test should be explicitly stated. Dates should be given for the collection of new data and the establishment of new norms. ESSENTIAL

B. Interpretation.

1. The test, the manual, record forms, and other accompanying material should assist users to make correct interpretations of the test results. ESSENTIAL

   1.4. If any systematic error resulting from testing conditions, regional factors, and other things, is likely to enter the test score, the manual should warn the user about it and discuss its probable size and direction. ESSENTIAL

2. The test manual should state implicitly the purposes and applications for which the test is recommended. ESSENTIAL

---

*The standards listed have been quoted from: APA, AERA, NCME Joint Committee, Standards for Educational and Psychological Tests and Manuals, Washington, D.C.: APA, 1966.

3. The test manual should indicate the qualifications required to administer the test and to interpret it properly. ESSENTIAL

C. Validity.

1. The manual should report the validity of the test for each type of inference for which it is recommended. If its validity for some suggested interpretation has not been investigated, that fact should be made clear. ESSENTIAL

2. Item-test correlations should not be presented in the manual as evidence of criterion-related validity, and they should be referred to as item-discrimination indices, not as item-validity coefficients. ESSENTIAL

3. If a test performance is to be interpreted as a sample of performance or a definition of performance in some universe of situations, the manual should indicate clearly what universe is represented and how adequate is the sampling. ESSENTIAL

   3.1 When experts have been asked to judge whether items are an appropriate sample of a universe or are correctly scored, the manual should describe the relevant professional experience and qualifications of the experts and the directions under which they made their judgments. VERY DESIRABLE

   3.2 In achievement tests of educational outcomes, the manual should report the classification system used for selecting items. DESIRABLE

7. If the author proposes to interpret the test as a measure of a theoretical variable (ability, trait, or attitude), the proposed interpretation should be fully stated. The interpretation of the theoretical construct should be distinguished from interpretations arising under other theories. ESSENTIAL

D. Reliability.

1. The test manual should be report evidence of reliability that permits the reader to judge whether scores are sufficiently dependable for the recommended uses of the test. If any of the necessary evidence has not been collected, the absence of such information should be noted. ESSENTIAL

   1.3 The standards for reliability should apply to every score, subscore, or combination of scores (such as a sum, difference, or quotient) which is recommended by the test manual (either explicitly or implicitly) for other than merely tentative or pilot use. ESSENTIAL

2. In the test manual reports on the reliability or error of measurement, procedures, and samples should be described sufficiently to permit a user to judge to what extent the evidence is applicable to the person and problems with which he is concerned. ESSENTIAL

3. Reports of reliability studies should ordinarily be expressed in the test manual in terms of variances for error components (or their square roots) or standard errors of measurement, or product-moment reliability coefficients. ESSENTIAL

4. If two forms of a test are published, both forms being intended for possible use with the same subjects, the means and variances of the two forms should be reported in the test manual, along with the coefficient of correlation between the two sets of scores. If necessary evidence is not provided, the test manual should warn the reader against assuming comparability. ESSENTIAL

E. Administration and Scoring.

1. The directions for administration should be presented in the test manual with sufficient clarity and emphasis that the test user can duplicate, and will be encouraged to duplicate, the administrative conditions under which the norms and data on reliability and validity were obtained. ESSENTIAL

2. The procedures for scoring the test should be presented in the test manual with a maximum of detail and clarity so as to reduce the likelihood of scoring error. ESSENTIAL

F. Scales and Norms.

1. Scales used for reporting scores should be so carefully described in the test manual as to increase the likelihood of accurate interpretation and the understanding of both the test interpreter and the subject. ESSENTIAL

    1.1 Standard scores should in general be used in preference to other derived scores. The system of standard scores should be consistent with the purposes for which the test is intended, and should be described in detail in the test manual. The reasons for choosing that scale in preference to other scales should also be made clear in the manual. VERY DESIRABLE

3. Local norms are more important for many uses of tests than are published norms. In such cases the test manual should suggest appropriate emphasis on local norms and describe methods for their calculation. VERY DESIRABLE

4. Norms should be reported in the test manual in terms of standard scores or percentile ranks which reflect the distribution of scores in an appropriate reference group or groups. ESSENTIAL

5. Norms presented in the test manual should refer to defined and clearly described populations. These populations should be the groups to whom users of the test will ordinarily wish to compare the persons tested. ESSENTIAL

APPENDIX C D

A Recommended Design for Measuring the Effect of Certain Factors on Occupational Competency Examination Performance

Factor A - A_1: Certificate of completion of vocational or technical 2 or 3-year secondary level program

A_2: Certificate as above in a post high school level program

A_3: No Certificate

Factor B - B_1: Apprentice level workers with from one to two years of experience

B_2: Journeyman level workers with up to 6 years of experience

B_3: Experienced journeymen level workers with from 7 to 10 years of experience

	B_1	B_2	B_3
A_1	n = 20 to 30 in each cell		
A_2	Criterion variables - part scores		
A_3			

Analysis of Variance Table*

Source of Variation	Degrees of Freedom
Between A_i	2
Between B_i	2
Between Parts of test (P)**	4
Interaction, A x B	4
Interaction, A x P	8
Interaction, B x P	8
Interaction, A x B x P	16
Between individuals within subgroup	261
Residual	1044
TOTAL	1349

* Assume n = 30 within each subgroup (A_iB_i)

** Assume 5 parts

References

1. Ahmann, J. S., and M. D. Glock, Evaluating Pupil Growth. Boston: Allyn and Bacon, 1963.

2. Anastasi, A. Psychological Testing, N.Y.: MacMillan, 1961.

3. Bloom, B. S., et. al. (Eds.). Taxonomy of Educational Objectives, Handbook I: Cognitive Domain, N.Y.: McKay, 1956.

4. Cronbach, L. J. Essentials of Psychological Testing, N.Y.: Harper, 1960.

5. Ebel, Robert L. "Obtaining and Reporting Evidence of Content Validity." Educational and Psychological Measurement, v. 16: 269-282, 1956.

6. Fattu, N.A. and D.M. Medley. Rationale and Construction of the k-Systems Performance Test. An unpublished report of the Air Research and Development Command, Lowry Air Force Base, Denver, Colorado, n.d.

7. Fleming, J. W. and E. K. Hankin. "College Credit for Occupational Competency." An unpublished report, Vocational and Adult Education, Florida State University, n.d.

8. Huddleston, Edith M. "Test Development on the Basis of Content Validity" Educational and Psychological Measurement, v.16: 283-293, 1956.

9. Kazanas, H.C. and L. D. Kieft. "An Experimental Project to Determine More Effective Vocational Teacher Certification Procedure in Michigan by Competency Examinations." Ypsilanti, Michigan: Eastern Michigan University Press, 1966.

10. Lennon, Roger T. "Assumptions Underlying the Use of Content Validity," Educational and Psychological Measurement, V. 16: 294-304, 1956.

11. Medley, Donald M. "The Effects of Item Heterogeneity and Guessing on the Accuracy of a Test Score." An unpublished report. City University of New York, 1962.

12. Micheels, W. J. and M. R. Karnes. Measuring Educational Achievement, N.Y.: McGraw-Hill, 1950.

13. Stokes, R. R., "The Split-Response Technique," Phi Delta Kappan, XLVII: 271-2 (January, 1966).

14. Thorndike, R. L., and E. Hagen. Measurement and Evaluation in Psychology and Education, N.Y.: Wiley, 1962.

15. The APA, AERA, NCME Joint Committee, Standards for Educational and Psychological Tests and Manuals, Washington, D.C.: APA, 1966.

APPENDIX E

CONSTRUCTING VALID OCCUPATIONAL COMPETENCY EXAMINATIONS[1]

by

Edward K. Hankin[2]

Without a doubt Dr. Impellitteri has prepared a thorough and scholarly paper on this topic. He clearly identifies its scope and provides guiding principles leading to his conclusions and clarifying their interpretations.

This reaction is divided into two major sections; one identifying points of agreement, and the other raising questions to either challenge his statements or to express disagreements.

In structure these sections will parallel the format of Dr. Impellitteri's presentation. This practice should facilitate the coordination of this reaction to his presentation.

I. Points of Agreement.

The three questions posed in the introduction adequately identify the purpose of the paper. Full answers to these questions should suffice for this aspect of establishing a nationwide occupational competency examination program.

The defining and illustration of validity and reliability are pertinent to the clarification of the meanings of the questions posed and provide an adequate basis for answering them. He indicates the essential characteristics of a test which contribute to these qualities and stresses their importance. His emphasis on this suggests the importance of these considerations for the testing program under discussion.

The author properly identifies three types of validity: content, construct, and criterion-related. Discussion of construct and content validity is quite sufficient for the purpose, providing a good foundation for the later discussion of procedures. More will be said in the second section of this paper regarding criterion-related validity.

In discussing the evidence of construct validity, the rationale suggesting significant positive correlation between the written and the performance parts of the test in a given occupation is quite logical and true. Other parts of this particular discussion will be referred to in Part Two.

One of the portions of Dr. Impellitteri's paper with which I am in high agreement is his discussion of the validation of part scores. Certainly in the construction of occupational competency examinations this should not be overlooked. Most of the occupations for which these examinations will be administered are quite broad in scope and high competency in one aspect of the occupation does not compensate for inadequate competency in another part. Probably no one who is well acquainted with the nature of skilled and technical occupations would take issue with much of what is stated.

---

[1] A Reaction to the paper of Dr. Joseph T. Impellitteri

[2] Dr. Hankin is Professor of Education at the Florida State University, Tallahassee, Florida

In this reaction some reference will be made in Section Two with reference to the intercorrelation of the part scores. This is the only portion of this discussion on the validation of part scores which I would challenge.

The author properly stresses the importance of norms as the basis for interpreting scores and establishing cut-offs. His argumentation clearly and adequately supports his contentions. Common practice in other fields of national testing such as NTE and GRE would favor the use of norms expressed in terms of standard deviation rather than in percentiles. The author has called attention to this in his footnote. Most persons who would be concerned with normative scores from these examinations would not be troubled by the matter of familiarity. As with other national testing programs, conversion tables could be provided for those whose statistical concepts are limited to centiles.

With reference to the group or groups to be used for establishing norms, the author raises significant questions which, as he indicates, must be answered after further study of the problems. Presumably one simple but unrevealing answer is that the norms should be established on the basis of people in the same occupation who are now teaching and whose competencies are satisfactory or on the basis of people employed in industry who are alike in competency to the people we want to have teaching. The identification of such groups for purposes of normalizing would be difficult but not impossible. These questions are somewhat related to questions regarding criterion related validity which will be discussed in Section II of this reaction. Certainly his "geographical factors" and "training factors" are important considerations.

This reactor is also in full agreement with the author's contentions regarding the necessity of norms for part scores. The whole purpose of having part scores would be defeated if norms were not provided for interpretation of raw scores. Hopefully the use of well established norms for part scores would compensate in part for less than perfect content-validity and construct validity. In effect it would superimpose the element of criterion validity, especially to the extent that the rain of improper or poorly constructed items would fall alike on the norm groups and the tested subjects.

The plan for constructing the examinations is quite logical and to the reactor appears to be almost complete. The matter of cost and practicality, which the author indicated he had not considered, might call for some scaling down of some of the quantitative steps. Questions regarding this matter will be raised in Part II of the reactions.

The steps as outlined would especially apply to the construction of a written examination and this is implied by some of his phrasing. Certainly the limitations of several of these steps in their application to performance testing should not go unnoticed.

Especially if the examinations are to be administered by trade and industrial education personnel at various times in many centers throughout the nation the test manual, as described in Step 8, is most essential.

The author's recommended procedures for content validity determination are consistent with his previous discussion of this quality and are quite adequate.

Similarly there is no real argument as to what he says about the construct validity as a general statement and assuming the readers of this paper are familiar with the processes required. It might be safe to assume that the technicians who would deal with this aspect of the validation procedures would be qualified in this respect.

Under the heading Empirical Keying the author describes procedures which are appropriate for this type of examination. Again this especially applies to the written examination. Further reference to this section will be made in Part II of the reactions.

No issue is taken with the author's statement regarding procedures for determining reliability of the examinations, providing two or more concurrent forms of examinations are produced. If the multiple forms are not developed, it would then be necessary to engage in one or the other of the alternative processes of determining reliability such as odd-even item correlations or a test-retest procedure. Probably the described procedures would best apply to the performance examination since the alternative procedures are less appropriate and the examination construction energies would be better directed to developing multiple forms. Certainly, as the author indicates, reliability must be determined for each part score portion of the total examination.

The discussion under item analysis of test results is quite satisfactory to the reactor, especially in terms of the written part of the examination. There might be considerable difficulty trying to apply the procedure to the performance examination, especially if there is only a limited number of tasks for each part score section. The procedures described are quite well-established by authorities in the field of paper and pencil testing.

For obvious reasons, no reaction is called for with reference to material included in the Appendix.

II. Questions to Challenge or To Express Disagreements.

This section of the reaction paper essentially calls attention to the ways in which the reactor differs with the author. It represents the reactor's judgments and opinions so he will attempt to explain why he differs and what he thinks ought to be.

For the most part the section is expressed in the third person without specifically identifying the reactor. Please keep in mind it is the reactor's expression, nonetheless.

In discussing the types of validity the author identifies three types but dismisses criterion-related validity. He explains that he does so because the competency examinations are not intended to be used as predictors as are aptitude tests. This seems to be an unnecessary and undesirable restriction applied to criterion-related validity.

Actually, later in the paper under the Plan for Constructing the Examinations, he describes procedures which in effect establish criterion-related validity. Steps 5, 6, and 7 on page 15, aimed at refining the items and establishing normative data, at the same time serve to establish criterion-related validity.

The criterion is the occupational competency of those to whom the tests are administered. Please note that the tests in each case are administered to persons with established occupational competency and not to prospective teachers whose competency is yet to be determined. The written test items to be eliminated are those for which satisfactory responses are not obtained from occupationally competent persons.

Since these processes are needed to develop and normalize the examinations, criterion-related validity is obtained without especially enlarging the task. This does not mean the tests are designed to be predictive, which seems to be the author's prime concern.

In discussing the question of validation of the written and performance examinations separately as opposed to considering them together the author suggests that the decision is irrelevant. The implication is that if the worker has the knowledge he has also the skill related to it, and vice versa. Furthermore, he implies that everything covered by written examinations will also be covered by performance examination. If these assumptions were true the performance examination would not be needed. The reactor contends that this is not true.

Some items of knowledge closely associated with performance will naturally be tested as part of the performance examination. When this is the case there is no need for them to also appear in the written examination. Furthermore there are some things which are more readily tested in the written examination and need not be tested in the performance examination.

These conditions suggest that neither the written examination nor the performance examination will have full content validity by themselves. Consequently, the content validity must be established for both parts together, as if they were one examination. In all probability neither part by itself could be completely valid so far as content is concerned, hence the content validity is improved by considering them together as a whole. It would be harmful to consider the two parts separately.

With reference to evidence of construct of validity the author discusses the degree of correlation which could be expected between scores on a written test and scores on a performance test in the same occupation. He states that if the coefficient of correlation approached the limit of +1.00 there would be no necessity for using both tests; they would be measuring the same thing. Actually this is an improper interpretation of the meaning of the coefficient of correlation in this situation. As Otis has pointed out, one proper interpretation of the coefficient of correlation is that it indexes the extent to which two different things are caused by the same third thing.

In the context of this subject the third thing is a combination of training, experience, and aptitude. Out of this combination grows the thing which is measured by the written examination and also the thing which is measured by a performance examination. To the extent that some individuals have inadequate training, experience and aptitude they will do poorly on both types of examinations. To the extent that they had ample training and experience and high aptitude they will do well on both parts. Some, because of differences in training, experience and aptitude, will do better on one part than the other. The proper inference is that the correlation will be positive and it might be high depending upon the group to whom the examination is administered for validation purposes.

If the examination is administered to a carefully selected group of well-qualified workers in an occupation it should be expected that the correlation between the performance and the written examination would be quite high; assuming the examinations were valid. In this situation with most of the examinees performing at a high level in both examinations the coeffieient of correlation would be reduced in magnitude because of the limited range in performance. The coefficient would approach +1.00 only if the written and performance examinations were administered to a group ranging from complete incompetents to full competency, with a minimum number with "mixed" competency.

Since the purpose of the examination is to identify a fully qualified person as compared to one who is only partially qualified we should expect a coefficient higher than +.60 between the written and the performance parts when they are administered to a large group representing a wide diversity of competency.

No reference was made in the first section to the discussion on the effect of guessing on reliability and validity. As the author states there is no consensus among authorities regarding correction for guessing.

If, in multiple choice items the distractors are well written only a complete novice will answer by guessing alone and his chances of guessing right are proportionate to the number of distractors. With a very large number of such items a novice is likely to obtain the low score he deserves without introducing a further penalty for guessing. On the other hand those who know the answers select the correct responses with no need to guess. The individuals who fall between these two extremes might, because of their possession of some knowledge in the realm of an item, eliminate some of the distractors and narrow their choices for guessing to two or three or four possible responses. The more they know the more distractors they eliminate and the higher their chances of guessing correctly. Thus, in a large number of items they will score higher than the novice and not so high as the expert, which is about where they belong. The nearer they are to the expert in their competence the more the guessing odds are in their favor and the nearer they are to the novice the greater the guessing odds are against them. Thus the range of uncorrected scores approximates the range of ability among those responding to a large number of items. This effect approximates the purpose of Stokes "split-response technique" and it is the reactor's suggestion that the matter of guessing be ignored both in scoring and in the use of corrections for guessing.

In the latter part of the author's discussion of the validation of part scores he suggests that if the parts have a high positive correlation the part scores would be useless. The observations as he makes are true for battery and sub-test type examinations in other fields, such as for intelligence and other aptitudes, for the intention there is to assess independent and largely unrelated characteristics of people. This is not the situation in occupational testing.

A fully qualified person should perform well in all parts of the competency examination. Conversely, a completely incompetent person will perform poorly in all parts of the examination. To the extent that an individual's qualifications were broad but limited, he would perform moderately in all parts of the examination. Only when individuals have had narrow experience in certain portions of an occupation will they perform well in some parts and poorly in others. (These results of course would be obtained only to the extent that the several parts were valid.)

On this basis it is here contended that the intercorrelations between the several parts of the test should be positive and high, though probably coefficients as high as +.95 could not be achieved. Such coefficients should exceed +.60. They certainly should not be so low as +.20 or +.30, assuming well-qualified persons were included in the population being tested. If such low and almost insignificant coefficients of correlations were obtained there would be good reason for questioning the validity of the examinations parts.

The suggestion that the sample for standardization should have at least 500 to 700 workers from a given occupation seems unreasonably and unnecessarily large. Aside from the fact that administering examinations to that many people for the purposes of obtaining normative data would be extremely expensive and time taking (which the author indicated he was not considering) a carefully selected smaller sample probably would serve the purposes better than such a large group. Even considering the regional variations, such norms ought to be obtainable from a sample of not over 100 workers and possibly as small as 50.

The time and expense required to identify a carefully selected representative small sample of the universe would probably be much less than the cost of extracting data from the suggested large sample. Furthermore, the results are likely to be of better quality.

In addition to the test manual called for in Step 8 there would also need to be prepared a booklet of information for those who contemplate having their occupational competency tested: the prospective teachers. This information booklet should give some indication of the scope and content of the parts of the examination so that the prospective examinee could refresh himself if needed or decide to not take the examination if he realizes his incompetence. Booklets should also describe the examination procedure and give details about registration and administration which would assist him in planning to take the examinations. This information would be something along the lines of the booklets provided for ETS's Graduate Record Examination, National Teacher's Examination and Teacher Education Examination Program.

In the discussion of empirical keying there is an implication which should be challenged. The item by item analysis displayed in Table 4 is for an examination intended for carpenters. With this purpose the comparison might better be made by administering the examination to a control group consisting of persons with characteristics similar to carpenters but different in their occupational field of training and experience or without any such qualification. Such a comparison would more clearly reveal the items which are most appropriate for examining carpenters than does the comparison shown.

In the discussion of this table, Items 1, 2, and 200 seem to be rejected for carpenters simply because the construction workers answered them almost as well or better. Contrary to this inference it is quite probable that carpenters should know things which other construction workers also should know: otherwise they would be unqualified as carpenters. Similarly, there are many things which other construction workers know which carpenters should also know, so this inference is viewed as improper.

In the discussion of the independence of part scores on page 18 the author again calls for low intercorrelation as an index of the usefulness of the part tests. As has been previously discussed, this is viewed as an improper conclusion. The part tests are intended to measure different portions of the same occupation. Well-qualified people should perform high in all parts. Novices could be expected to perform low in all parts. Some individuals with narrow but intensive experience in certain parts of the occupation will perform much better on some parts than on others. To the extent that these are in the minority the intercorreltations for a population ranging from novices to experts should produce rather high coefficients, probably above +.60.

In this section the author also suggests factor analysis as an empirical measure of utility of part scores. What he is getting at is that the subtests producing part scores ought to each be measuring a different aspect of the occupation. This is a matter which can be most readily accomplished in the test construction where the content of the occupation is being sampled. An initial step would be to subdivide the content into several mutually exclusive divisions and design a subtest for each division. A particular item of content should not be tested in more than one subtest. The suggested factor analysis procedure would lead to improper conclusions to the extent that it identified test items on two completely different elements of content which were answered equally well by many of the examinees. It is quite probable that this would be a desirable rather than an undesirable characteristic.

In the authors discussion of factors related to scores on competency examinations, he raises a number of questions particularly concerning the population to be used for establishing norms. He is correct in suggesting that the answers to his questions are crucial but he does not pose the answers.

It seems appropriate to the reactor that the population used for normalizing these examinations should be made-up of a carefully constructed sample of the universe representing the full range of competency from the novice to the recognized fully trained and experienced worker in the occupation. This sample probably would include some beginning apprentices or beginning students in an occupational preparatory curriculum together with some highly proficient experienced workers in the occupation. In all probability it could also include some who had finished their training but had no experience, such as those completing apprenticeship or completing their in-school preparation. Workers with too many years of experience in the occupation might properly be excluded, first because their responses might suffer because of their age and secondly because they are far removed by time from their training and might not be up-to-date in their occupational competencies. The criteria for selecting the normalizing population probably should establish an upper age limit as well as a maximum number of years of experience, such as five or six years following the period of training. Such criteria would not unduly restrict the potential size of the sample and they would eliminate some of these contaminating elements.

APPENDIX F

A LIMITED FIELD TEST OF THE AUTOMOTIVE COMPETENCY EXAMINATION

by

Ray A. LaBounty[1]

Need for the Study.

The shortage of well qualified vocational and technical education teachers is increasing to the point of becoming critical in many cities of the nation. In the Detroit Public School System, a number of school shops have had to be closed due to a lack of qualified vocational teachers. (Detroit Public Schools, 1965). Similarly, in New York City, several hundred vocational teachers are needed for newly developing and expanding programs in vocational education. (Shapiro, 1965). Similar situations exist in many other school systems throughout the United States. A recent study of this problem in Michigan (Department of Public Instruction, 1965) has revealed that there are 291 vocational and technical teachers needed for this year to fill the vacancies in the state. This study further points out that there will be a shortage of 634 teachers by 1966 and 1,862 teachers by 1970 due to the anticipated expansion of vocational and technical programs in the state.

At the very time when the need for more and better programs in vocational and technical education is essential, the programs cannot be developed or expanded because of this damaging shortage of qualified teachers.

It is a fact that the shortage will become even more critical in the coming few decades as more money becomes available for vocational and technical programs and as changes in the world of work increase at an accelerated rate. As Smith (1963:44) emphasized, "One of the most stubborn problems to be met in the expansion of vocational education is the limited supply of competent teachers."

To overcome such a critical and expanding problem, new horizons should be opened for the recruitment, selection, preparation, supply, and certification of well qualified vocational and technical education teachers.

> Just as there is a need for a more comprehensive program for the preparation of individuals to enter the labor force, so it follows that the programs of preparation for vocational teachers must be more rigorous and often quite different from those now provided. (Swanson and Kramer, 1965:170).

The effectiveness of all vocational and technical programs depend upon the adequate preparation, supply, and certification of teachers. Vander Werf (1965:408) has stated:

> . . . A major key to the effectiveness of learning in vocational programs in the next few decades will be the recruitment, selection, and preparation of teachers.

---

[1] Mr. LaBounty is Professor and Head, Department of Industrial Education and Applied Arts, Eastern Michigan University, Ypsilanti, Michigan.

> In the future, many thousands will be needed at the very stages when positions elsewhere requiring similar competencies will be more attractive financially. Recruitment programs will have to be imaginative and distinctive, backed up by programs of preparation worthy of the enticements.

It is clear, therefore, that "solutions to the teacher-training problem must be found, for the quality of vocational programs is determined in large measure by the quality of instruction." (Office of Education, 1963:11).

The problem of vocational and technical education teacher shortage has several logical explanations:

1. It appears to be related in part to the low salaries paid to teachers in comparison with the salaries paid to workers in other occupations, many of which require less post high school preparation for entrance into the occupation than that required for teachers.

2. Related to this is the fact that, on a national basis, teaching as a profession, is rated below a large number of other professions on prestige status scales. This becomes even more complicated for vocational and technical education teachers, due to the attitudes held by many groups in our society that "work" involving manual skills and direct involvement with tools and machinery is beneath the "dignity" of a person with a college education. (Smith 1963:44).

The problem of acquiring competent vocational teachers is aggravated by the traditions and standards that have developed in the teaching profession. Rank, prestige, status, salary scales, and certification requirements are geared to years of schooling, degrees obtained, and seniority. Furthermore, the salary scale for teachers may not compare favorably with journeyman's pay; and teachers' pay raises may be smaller and more infrequent. Why should any skilled journeyman shift to full-time teaching under these circumstances? These attitudes tend to dissuade those persons who might have an inclination toward vocational and technical teaching.

3. The present certification procedures and requirements of most states can be considered unrealistic. Most states require that the prospective vocational teacher have a Baccalaureate degree plus at least three years of occupational experience. This amounts to seven or eight years of preparation for vocational teaching. This tends to discourage many young and ambitious people from entering the profession.

4. The sources and methods of recruitment, selection, and certification have been inadequate. At present, most vocational teachers are recruited only from one source, industry, without any attempt to find other favorable sources or methods of recruitment and certification. (Willis, 1963).

These and several other factors have created, and will continue to create, the critical shortage of vocational and technical education teachers unless new approaches are developed to alter the situation.

Colleges and universities have for years graduated teachers in vocational agriculture, home economics and distributive education who were certified by the state to teach these vocational subjects on the basis of college preparation. For trade and industrial education teachers, however, the situation has been quite different. It has been rather common practice to recruit trade and industrial teachers from industry and/or business. Journeymen, who were willing to teach, were provided a number of "professional education" courses and a special certificate were sent out into the classroom to teach--with very little consideration accorded to whether they were "qualified" to teach on any other basis than vocational experience. Such a plan for staffing vocational education programs has not kept pace with the demands placed on trade and industrial education. The preparation and the certification procedures must be reviewed and revised, or changed, in accordance with the new philosophy and practices of vocational education as it has been expressed in the new vocational legislation.

The new Michigan State Plan for Vocational Education indicates that the prospective trade and industrial education teacher shall possess a Baccalaureate Degree plus three years of occupational experience in the occupational areas concerned. Under certain circumstances, the candidate may be administered a competency examination. This last provision, although a potential source for teacher certification, has not been adequately explored in Michigan. There are many individuals in Michigan or elsewhere in the field of industrial education who are capable of teaching vocational-industrial subjects either because of their intensive formal education in vocational-technical and/or engineering programs, or because of their prolonged and varied teaching experience plus a limited amount of work experience. Though competent to teach vocational subjects, the present system of certification makes it most difficult for such persons to become instructors in reimbursible vocational programs. A new approach concerned with certification of trade and industrial teachers that will either replace or supplement the present situation, is needed if the quantity and quality of trade and industrial teachers are to meet vocational education demands.

The approach that seems to be most promising at the present is to organize, expand, and standardize the provision stated in the Michigan State Plan for Vocational Education: to certify trade and industrial teachers by competency examinations. This study was initiated to determine a basis for certifying teachers under this provision. The purpose of the study was to investigate this matter and to present some recommendations determined from presently used competency examinations in the United States; and to prepare the necessary competency examinations and testing procedures to be used in the state of Michigan.

Objectives of the Study.

The general hypothesis of this study has been that better results can be obtained in trade and industrial teacher certification in Michigan. This can be accomplished by developing and using well designed competency tests and testing procedures which will be accessible to more majors in industrial education who are interested in becoming vocational teachers, but who do not meet the existing state requirements. These tests will also be available to other individuals with various backgrounds who desire to but could not obtain vocational certification because of existing work experience requirements. This new approach can be carried out by initiating a cooperative short term project with the Michigan State Department of Education, Division of Vocational Education, and Eastern Michigan University to study, develop, and refine such competency tests and testing procedures.

The general objectives of the study are to:

1. Review the literature, particularly the new State Plans of Vocational Education of the various states and territories to determine the present practices and requirements in trade and industrial teacher certification.
2. Determine, develop, and refine the necessary testing trade and industry instruments and testing procedures to be used for teacher certification.
3. Provide the basis for a continual revision, reviewing, exploring, and evaluation of the certification procedures in Michigan.
4. Make the results available to all teacher education institutions in Michigan.
5. Evaluate whether such an approach is practical and will produce desired results.

Definitions of Terms

COMPETENCY EXAMINATION (TRADE TEST)

A test or examination including three parts; written, performance and oral; designed to determine a level of technical knowledge and skills of a teacher candidate in a particular trade and industrial area.

RECOGNIZED WORK EXPERIENCE

Formal full-time and/or part-time employment undertaken by a teacher candidate for a specified length of time in a specific occupation considered by the State Board of Vocational Education as being necessary in obtaining technical trade knowledge and skills.

STATE PLAN

An agreement between a state board for vocational education and the U.S. Office of Education describing (a) the vocational education program developed by the state to meet its own purposes and conditions, and (b) the conditions under which the state will use Federal vocational education funds (such conditions must conform to the Federal acts and the official policies of the U.S. Office of Education before programs may be reimbursed from Federal funds. (American Vocational Association:17).

TEACHER CANDIDATE

An individual who is planning to become vocationally certified to teach trade and industrial courses in specific trade areas.

TEACHER EDUCATOR

A vocationally qualified professional person responsible for the preparation and in-service training of teachers. He assists teachers or prospective teachers in securing the professional knowledge, ability, understanding, and appreciation which will enable them to meet certification requirements or advance in teaching positions. (American Vocational Association:19).

TRADE AREAS

A group of industrial occupations (usually apprenticeable) which require a high degree of skill, technical knowledge, and mechanical training and dexterity; usually in a wide range of related activities and secured through a combination of job instruction and work experience. This is exclusive of agriculture and business.

TRADE ANALYSIS

The procedure of breaking down a trade or occupation to determine the teachable content in terms of operations, tools, processes, and technical information to be organized into a course of study and arranged according to a sequence of difficulty.

TRADE AND INDUSTRIAL TEACHER

Any teacher that has been certified by the State Board of Vocational Education as being qualified to teach trade and industrial courses.

VOCATIONAL AND TECHNICAL EDUCATION

Training intended to prepare an individual to earn a living in an occupation in which success is dependent largely upon technical information and an understanding of laws of science and technology as applied to modern design, production, distribution and services. (American Vocational Association:22).

TRADE AND INDUSTRIAL EDUCATION (VOCATIONAL EDUCATION)

Instruction which is planned to develop basic manipulative skills, safety judgment, technical knowledge, and related occupational information for the purpose of fitting persons for initial employment in industrial occupations and upgrading or retraining workers employed in industry. (American Vocational Association:20).

VOCATIONAL CERTIFICATION

The approval action, based on minimum standards adopted in the state, taken by legally authorized school authorities on the professional and technical qualifications of teachers. (American Vocational Association:19).

Summary of State Directors' Comments

Those state directors who disagreed that competency exams could serve as a substitute for required work experience presented their comments as to how competency exams could be used. Essentially, they believed that competency examinations should never be used as the sole means of certification but rather they should be used to verify occupational experiences, knowledge, and proficiency skills; to act as a partial substitute for work experience; and to furnish additional information for evaluation. Actual work experience was considered to be very important in the preparation and certification of vocational teachers.

Several state directors felt that there was no possibility of using competency examinations as a means for certification. If these exams were used at all, it would be only in a manner where the results of the examination supplemented the entire requirements established by the state. These viewpoints expressed doubt that competency examinations could ever be developed which would be able to effectively evaluate judgments, occupational competency, complete understandings and special skills which are obtained by actual work experience.

Additional comments brought about a variety of viewpoints. One comment which expressed the view of many of the state directors is that "Competency or trade tests can be a part of a balanced certification program and may be used to assist in determining the extent and quality of trade knowledge in both theory and practice." However, some work experience is still considered important to gain the necessary trade knowledge and skills. No other specific procedures were mentioned where such knowledge and experience might be obtained.

Development of Competency Examinations and Testing Procedures for the State of Michigan

It was determined from the evaluation of the survey questionnaire, that it was necessary to develop three types of tests. These were oral, written (objective), and performance tests for each subject area. The written portion of the instrument was found to be the most difficult and time consuming part to develop. The step-by-step procedures taken in the development of the written portion of the testing instruments are as follows:

1. An analysis of the trade area was first developed. Several different analyses for each trade area were assembled, analyzed, and evaluated to provide information concerning the trade area. A comprehensive analysis of the trade area was then derived from the several different analyses and reviewed by the research staff and by an Eastern Michigan University vocational teacher educator concerned with that particular trade area.

2. A large number of multiple choice and "True and False" questions were developed and organized into groups coinciding with the various parts of the trade area analysis. These questions were selected either from presently used tests or were developed from subject matter in that trade area. The number of questions selected for each group approximated the percentage of emphasis that was given each part of the trade analysis. After a sufficient number of questions had been selected (approximately 1500 questions for each trade area), an initial screening of these questions was made by the research personnel and the Eastern Michigan University Vocational teacher educator concerned with that particular trade area. The questions were then typed in a rough draft.

3. A committee for each trade area was selected to evaluate the trade analysis and the rough draft of the trade examination. The committee consisted of two members representing industry in the particular area involved; two members representing vocational education, the Eastern Michigan University vocational

teacher educator acting as chairman, and the research staff of this study. The purpose of each committee was to evaluate the trade analysis as to its completeness in trade area coverage and to the degree of importance that each area of the trade analysis was given on the tests. Next, the committee evaluated the first draft of the tests; not their design or length, but the technical content of the tests. Certain areas were given too much emphasis by way of the number of questions involved, while other areas needed more emphasis. Certain questions needed to be revised for clarity and correctness. After the committee members individually reviewed the trade analysis and the tests, they assembled at Eastern Michigan University campus to discuss the tests and present their comments and recommendations. The committee also presented recommendations concerning the most common experiences which vocational trade teachers should be capable of performing in the area concerned. From these recommendations a suitable list of experiences was developed for use in the performance test for that area.

The semi-final revision was then completed on the written tests. The tests were then broken down into forms A and B consisting of about three hundred questions. Several questions (those considered to be basic) were included on both forms.

Field Test of the Automotive Teacher Competency Examination for the Second Rutger's Conference

To gather additional information on the feasibility of area or nation wide usage of competency examinations a very modest and limited field test was attempted.

Fifteen of the automotive competency examinations were administered in each of three states, Michigan, Kansas, and North Carolina.

The states were selected because of interest expressed by representatives at the first Rutger's Conference. The automotive area was selected because of its broad geographic appeal and the obvious identification of common subject matter content.

The automotive competency examination has not been revised since its development. It has not, in fact, been used in more than a very few instances prior to this field application.

Candidates taking the examination were asked to make comments on test questions as they completed the examination. It is believed that some revision may result from such criticism.

An attempt was made to gather certain information of those taking the examination without jeopardizing the anonymity of the volunteer. Each volunteer will receive copies of this paper, and by identifying his number he may compare results. No other information was collected.

Results indicated little correlation between test scores and age, years of teaching, or years of work experience.

No attempt has been made to treat the information statistically. To do so would be an exercise in futility in that we are not sure enough of the reliability of the data.

In one state the correlation between test score and years teaching was .12, while the correlation between test scores and work experience was .11. Such figures are hardly significant from which to draw conclusions.

Conclusions.

The real value to the test administration is in the eagerness by which it was received, both by directors and volunteers. The comments made on the score sheets were positive in tone and reflect a feeling of need for some measuring device by which we can begin to evaluate teacher candidates.

A most revealing feature is the very small number of people who have had college work in automotive mechanics. It would seem to follow that if we are to teach the technology, the teachers would have access to the content area in their professional preparation.

The volunteers were not requested to indicate the amount of college work they had completed. From the small number who volunteered the information it would appear that there are many automotive teachers in the sampling who do not hold the baccalaureate degree.

APPENDIX G

A LIMITED FIELD TEST OF THE AUTOMOTIVE COMPETENCY EXAMINATION[1]

by

Donn Billings[2]

Mr. LaBounty has done an excellent job of summarizing the vocational and technical teacher shortage throughout the country and the problems involved in the trade competency examination development. I would like to point out, however, that the description of the situation in New York City was perhaps misstated by the source that he quoted.

The new Michigan State Plan for vocational education and the uses of the trade competency examinations was very thoroughly outlined. It is interesting to note that even within the state it is difficult to get the various institutions, responsible for trade competency test development, to work together and to pool their resources.

Mr. LaBounty is to be commended for his attempt at providing a field test of an automotive teacher competency examination. However, I am not sure exactly what was learned from having 15 of these examinations given in Kansas and North Carolina along with those administered in Michigan. Perhaps the most important aspect of this was the fact that the states were willing to cooperate in this project.

The major problem to overcome, in the development of trade competency examinations on a national basis, is the reluctance of the various states to: (1) Share developmental projects, (2) Accept at par value other state philosophies and standards, (3) Exchange test instruments on an unlimited basis, and (4) Discharge traditional patterns of trade testing.

My personal concerns about any testing program on a national basis are: (1) The effectiveness of the measurements, (2) Security in the test procedure, and (3) The mechanics and problems of the operation and administration of this program. This program requires research in depth rather than simply pooling the job analyses and test items that have been developed in some of the other states. Unless these tests are developed and refined in a sophisticated form there will be no improvement over the present system. In addition to the traditional measurement of the trade competency of the individual, I would like to see research done and tests developed to measure, if possible, the individual's ability to teach or to develop into an effective teacher.

In closing, I would like to restate my plea for an emphasis on experimental research rather than on the traditional test development as a beginning for this type of a program. Based upon this research one could demonstrate the administrative mechanics that would be used for a national system of trade competency examinations.

---

[1] A Reaction to the paper of Mr. Ray A. LaBounty.

[2] Dr. Billings is Coordinator of Industrial-Technical Education, New York State Education Department.

APPENDIX H

PREPARATION, ADMINISTRATION, AND IMPLEMENTATION OF TRADE COMPETENCY EXAMINATIONS FOR COLLEGE-UNIVERSITY CREDIT

by

Joe L. Reed[1]

The Problem

In this age of astro-physics, astronauts, outer space, automation, numerical control, atomic energy and technological change, the need for technical personnel and especially the need for vocational technical teachers, is very much on the increase. Recognizing the challenge that must inevitably come with change, we are faced with two major questions, (1) who are or what are these vocational technical teachers, (2) where are we going to secure them?

The Importance of the Problem

National leaders have stated that our failure to fit American youth for the many technical jobs that exist for them is a national tragedy. We are graduating or terminating American youth from our schools in ever increasing numbers on the one hand, while jobs for which they are qualified to secure and hold are diminishing inversely proportionate to the demand for these jobs.

It is not now, nor has it ever been the philosophy of practical arts vocational educators that the general and cultural aspects of education should be curtailed or de-emphasized in any way for any student or group of students. No area of special vocational preparation can be any stronger than the foundation of general education upon which it must be built.

The time is long overdue when we in education should recognize that education should not only develop the individual in how to live the good life, but also how to earn an adequate income in order to afford the good life. Dr. James Umstadt, of the University of Texas, states that all too long the secondary schools have been producing students who are all dressed up academically with no place to go vocationally in the area of securing and holding worthwhile jobs.

It seems to be a reasonable assumption that the bridge that could span the gap between the expanding multitude of unemployed and underemployed youth and the multiplicity of increasing industrial-technical jobs that are going begging for lack of qualified workers is education and training. Supporting this assumption is a statement of the Ten Imperative Needs of Youth, developed by the National Association of Secondary School Principals in 1958 which states the first of the Imperative Needs of Youth is that of the development of saleable skills. This is not an ambigious vague statement or term. Saleable skills are those abilities which employers will employ and pay a salary or wage for services rendered.

---

[1] Mr. Joe L. Reed is Professor and Head of the Department of Industrial Education, The University of Tennessee, Knoxville, Tennessee.

A second assumption is one that is supported by empirical evidence gained through time and experience. It emphasizes that no educational program can be any stronger than the proficiency of the instructor in content and teaching techniques. This is especially significant when we conclude that employment opportunities are greatest in occupations requiring advanced levels of both general and special education in manipulative skills and technical knowledge. This conclusion precludes that a greater degree of competency will be required of the instructor in the future compared to the present or past.

The funding of the 1963 Vocational Education and other Federally aided acts, created many supervisory and administrative jobs that were filled by trade and industrial shops and laboratories teachers. This exodus of trade teachers further compounds the problem of teacher shortage.

Unlike many general education programs, trade and industrial teachers must instruct students, not only to the level of understanding and knowledge about something, but also to the level of ability to do something plus an appreciation of the importance of the job or operation. Because the primary purpose of vocational education is to prepare individuals for employment or advancement in an occupation, the instructional program is based on the requirements and practices of an occupation, and teachers must be equipped by practical experience and professional training to provide students with the occupational skills, knowledge, attitudes, and appreciations they need to fulfill job requirements.

According to Benjamin Franklin, a person can no more teach something he doesn't know than he can come back from some place he has not been. In order to demonstrate the skills of an occupation, as well as to explain the technical related knowledge of the occupation, a person must be a master of the occupation or trade.

The prime supporting pilasters in the educational bridge that spans the gap between unemployment and gainful productivity are highly skilled and knowledgeable teachers.

Unlike other vocational fields who secure their instructors from college campuses, the trade and industrial program must secure teachers from other sources.

Since most colleges and universities are not tooled up to develop the manipulative competencies and technical theory proficiencies that are required of trade and industrial teachers, these skills and knowledges must be secured through training and experience gained off the campus. The purpose of this paper is not to identify the precise way or ways in which these proficiencies and competencies must be gained. It is an attempt to suggest a possible plan for the organization, administration and implementation of a national program for the preparation and use of trade proficiency examinations to substantiate the mastery of competencies needed by trade and industrial teachers.

Need for the Examinations

In addition to the verification of occupational competencies, these examinations could serve many important purposes. A study by Earhart, of vocational technical training, and certification in trades and industry in the various states and territories, reveals that from 2 to 8 years of trade experience, beyond the apprenticeship or learning period was required for certification of instructors.

One of the uses of these examinations would be that of measuring the extent of skills and knowledge gained through these experiences to insure that it is reasonably uniform from one area of the nation to another and that it does represent 2 to 8 years comprehensive coverage of the occupation to be taught and not just a repeat of one year's experience 8 times. Another use could be that of verification of mastery of subject to be taught. Unfortunately, some of our employing academic administrators have a "sheepskin psychosis" that makes it difficult for them to understand that a person who has less than a Baccalaureate Degree may be a satisfactory or even an excellent teacher.

With the passage of the 1963 Vocation Education Act and the inauguration of many new and different types of programs, many individuals have been employed for teaching subjects who are not fully qualified according to present day standards. Trade competency examinations would serve to identify these individuals and prevent them from migrating to the regular trade and industrial program when we again return to normalcy.

Still another use of the examinations could be to discourage the employment of substandard teachers. They would serve as documentary evidence that these individuals are not fully competent if local and other school administrative units insist on employing them for pressure or emergency reasons.

One of the most important uses of trade competency examinations would be that of granting college-university credit for experience gained in industry. It would no doubt, in many cases, serve as an incentive to those in industry to come in to teaching since they may be given this all important gift of time in the form of college credits toward a Baccalaureate Degree. It would also serve as an incentive for those who come in to teaching to further their professional improvement by taking courses to gain more college credit.

Use of Examinations

Among the potential uses of national competency examinations are:

A. To provide state certification boards with an alternative to the "years of experience" requirement. This would provide useful information that state boards might be willing to consider; or may stimulate research to determine the amount of industrial experience required for certification.

B. To give university credit for work experience or experience gained in co-operative programs. It was considered important that valid instruments be available on which to base decisions where credit was involved.

C. To help raise salaries and prestige of vocational education, maintain high standards and to help teachers recognize important facets of the trade to emphasize in teaching.

D. To validate vocational teachers' competencies in the eyes of academic administration.

E. For teacher certification purposes, as evidence of competency, for reciprocity purposes between centers and states.

F. For teacher recruitment and selection.

   1. To screen individuals for pre-service and to plan for training of teachers.

   2. To identify vocational graduates who may make future teachers.

G. To identify sub-marginal and non-competent teachers who have been approved for teaching subjects in areas outside of their areas of preparation and experience.

   1. To deny renewal of a teacher's credentials when proven incompetent.

   2. To prevent non-competent teachers in special programs from migrating to regular trade and industrial programs in the future.

   3. As a diagnostic measurement for recommending more study or more work experience to meet future certification.

   4. To substantiate incompetencies of those who are certified through political pressures or special arrangements.

There was some feeling that the cost, time and trouble involved in taking proficiency examinations might discourage promising applicants. There was also recognition that by linking certification and college credit some craftsmen and teachers might be encouraged to embark on studies leading to a degree.

Definition of Terms

According to Melvin V. Keil and John W. Neubauer, in the 1965, April issue, Phi Delta Kappan, the following definitions were given:

> "Technical Education should be considered education directed toward an occupation in which success is dependent largely upon technical information and understandings of the laws of science and technology as they are applied to modern design, distribution and service. The student prepared for work through this curriculum must have a definite facility in mathematics and communications, skills which includes the ability to interpret, analyze and transmit facts and ideas graphically and orally. The technician is the link between the engineer and the skilled trades worker."

Vocational Education as a general term means education for work, any kind of work. Seen in this light, medical and legal education as well as business education and homemaking courses in high school are actually vocational education; but in industrial education the term is usually confined to job training at high school level, and as such, should be considered a form of education designed to develop skills and abilities encompassing knowledge and information needed by workers to enter and make progress in employment on a useful and productive basis. In other words, vocational education is organized instruction below college level to prepare the learner for a particular occupation.

Industrial Arts is industrial shop work of a non-vocational type which provides general education experience centered around the industrial and technical aspects of life and offers orientation in the area of appreciateion, production, consumption and recreation through actual experience with materials and goods. It also serves as an exploratory experience which is helpful in making occupational choices. Industrial Arts serves well as the practical application of mathematics and science as well as other liberal arts and its scope is wide enough to include the slow learner and the gifted student.

Industrial education is a generic, all encompassing term used to describe various types of education having to do with the production of material goods, including industrial arts, trade education and technical education. According to rules and regulations on the interpretation of Bulletin I, an industrial pursuit may be any of the following:

(a) Any industrial pursuit, skilled or semi-skilled trade, craft, or occupation which directly functions in the designing, producing, processing, assembling, maintaining, servicing, or repairing of any product or commodity.

(b) Other occupations which are usually considered technical and in which workers such as nurses, laboratory assistants, draftsmen, and technicians, are employed and which are not classified as agricultural, distribution and other business, professional or homemaking.

(c) Service occupations which are trade and industrial in nature.

Legally, the examinations could be given to individuals in any of these trade and industrial or other highly technical service occupations, including draftsmen, technicians, and nurses; however for the purpose of this presentation, reference will be made largely to the regular trade and industrial occupations or skilled trades.

Justification for College Credit

The occupational experience of a vocational teacher represents "his teaching field" preparation. In this report it corresponds to the mathematics which the mathematics teacher studies, the Spanish which the Spanish teacher studies, or the sciences which are included in the college curriculum for the training of a science teacher. The abilities which a vocational teacher has acquired and demonstrated through being employed as a senior worker (journeyman) in his occupation provides him with the content background for his teaching qualifications.

His preparation should also include professional courses which are concerned with the purposes, planning, presentation, and evaluation of instruction, plus advanced general education.

Recognizing the close parallel between the industrial vocational teacher's occupational experience and the subject teaching field content preparation, this program proposes to allow college credit toward a bachelor of science degree in education for industrial vocational teachers.

It is the underlying philosophy of this program that the college credit is being given for the skills and knowledge which are possessed by an experienced master of a skilled occupation. In the trades and crafts, this means a person with at least two years or more beyond the learning period of journeyman level experience. The occupational competency examination merely verifies the scope and quality of abilities developed through documented employment experience.

It is not the intent of the writer in this paper to suggest in any way that trade competency examinations should be designed and given in lieu of occupational proficiency. The intent and purpose of such a plan would be that of the documentation of occupational competency and proficiencies that are needed without regards to exactly how they are gained.

Plan for Implementation

On the basis of past experience and the experience of others with this type of program, it is recommended that three examinations should probably be given in the occupational field in which the applicant possesses a mastery of skill and knowledge. The examination should consist of the following:

1. A written examination (a minimum of 3 hours) shall be related to the technical trade or occupational information. It should include the sciences, mathematics, technology, print reading and job planning in the occupation.

2. A manipulative examination (a maximum of 6 hours) shall consist of performance of trade or occupational operations and job. This examination should be administered with actual machines, tools, materials that the individual would be working with in the trade or occupation.

3. An oral examination (2 hours) shall consist of an evaluation of trade or occupational knowledge including subjects; and personal qualifications.

It has been noted that some states who have a plan for competency examinations have dropped the oral phase of the examination. In spite of this there is reason to believe that much can be evaluated and measured about a prospective teacher as a result of this type of interview or examination. For example, a person may be very knowledgeable and skillful but still unable to communicate orally in correct, or reasonable correct, grammar.

Also there may be a determination made as to the general attitude and outlook of such a person. By proper type of questioning, it could be determined if a person is consistent in his thinking or if he is inclined to agree with leading types of questions, hoping to please the interrogator, judges or committee who are administering the examination.

The writer would strongly recommend that oral examinations be considered in a national plan for giving college credit for occupational competency. In many respects this examination is a pre-determination if the college or university would care ultimately to put a stamp of approval in the form of a degree on this person.

Provisions for Recording Credit

Regardless of the number of examinations or the amount of credit given, there must be some provision for recording credit or getting it on to the individual's college or university record. The best plan for doing this seems to be to secure approval through regular channels of the college or university of certain courses by title and number with the prescribed number of college credits to be given in each. In some cases it will be stipulated that credit in these courses may be earned through proficiency examinations only. A suggested plan is that in a course such as Industrial Education 3010, Related Science, Math and Technology in Occupations, 15 quarter hours of credit may be given. Industrial Education 3020, Manipulative Skills in Occupations, 15 quarter hours may be given. Industrial Education 3030, Knowledge of Related Subjects in Occupations and Personal Qualifications, 15 quarter hours of credit may be given.

This would provide for the granting of 45 quarter hours or one year of college credit for occupational competency in some technical trade or occupation. This procedure is recommended since there seems to be no academic way of recording credit by just writing out one year of college credit for trade experience. Credit must be recorded in the form of approved established courses in the curriculum regardless of whether they are completed for college credit or for credit earned through examinations. The simplest plan for recording credit seems to be according to types of examinations such as written, oral and manipulative rather than various units of an occupation since many occupations do not lend themselves to a division of 5 or 6 units or areas of instruction.

Cost of Proficiency Examinations

Many colleges and universities already have plans for granting credit through proficiency examinations. A typical entry in the college catalog may be as follows:

> "A proficiency examination may be given to qualified students in any academic course offered in the university on the recommendation of the head of the department and the payment of an examination fee".

At the University of Tennessee, the cost of the examinations for Industrial Education 3010 - 20 - 30 is $10 each or a total of $30 for the total credit that may be earned.

It should be understood by the student that through these examinations he may earn 45 quarter hours of credit, 30 quarter hours of credit, 15 quarter hours of credit or no credit, depending on the number of examinations satisfactorily passed.

The examinations may be given in the applicant's home town or in nearby major centers throughout the state, depending on the location and adequacy of equipment for administering the examinations. In some cases applicants, or students are charged fees for committee members' travel expense in administering the examination. In most cases students are not required or asked to pay in excess of $20 above the fees for the cost of the examinations.

It is suggested that qualified applicants be permitted to complete forms requesting the examinations at any time. Because of difficulty in refunding fees, it is suggested that fees should not be collected until the examination is announced.

This undergraduate credit should be applicable only toward a bachelor of science degree. It is not intended for, and should not be used in lieu of, any methodology certification courses that are required for certificate renewal.

Examination Committee or Team

It is recommended that these examinations be administered by a committee or team consisting of at least 4 to 6 individuals. Two of these individuals should represent the craft or occupation in which the trade examination is being given. They should be highly competent in skills in all facets of the trade. One might represent management, with the other representing labor. Since the earned credit may be used for certification purposes, a representative of the State Department of Education should be a member of the examination committee. The fourth person should be a representative of the college or university in which the credit is to be recorded and should be from the Industrial Education Department. Because of administrative duties, it may not be possible to always secure the service of the fifth and sixth members of the examination team. However, an opportunity should probably be extended to the Dean of Admissions since credit is to be registered officially by him for the examination. The Dean of the College of Education should also be invited to sit in on the examination since this is the school or college in which credit is to be recorded for the examinations.

Types of Examinations

In keeping with the latest recommended practices in test and measurements, it is recommended that the written examination should be of the objective type, preferably multiple choice questions. There should not be less than 3 choices on each question and not more than 5 choices. To insure maximum thinking by the teacher being tested, it is recommended that the questions be worded in such a way that the instructor is looking for the correct answer among the choices. On other questions the instructor would be looking for the exceptions which would be considered the correct answer. The multiple choice type of question seems to have many advantages over other short answer type questions; however, true-false, matching terms, completion and even a few essay or explanation type questions may be included if the writer feels strongly as to the use of them.

At the time the test is validated by an industrial committee, a key listing the right answer should also be prepared and validated at that time. The time of the examination committee should not be consumed in considering which are the right and which are the wrong answers. Actually the written part of the test could be scored by a graduate student or some individual who is capable of checking the key against the responses.

To assist the examination committee, a list of suggested questions should be prepared for the oral portion of the examination. These questions may be of two or more classifications. The first should be general questions such as, "Tell us something about yourself as to your background, trade experience, and education." This type question will make it easy for the person being examined to respond. The second list of questions should be specific or technical, starting with such statements or words as, what, where, when, how and why of certain things in the trade. These questions may not be followed exactly with every person being examined; however, they will provide committee members with at least basic questions that they can ask until other questions come to mind. In addition to testing the individual's technical knowledge of the trade or occupation, oral questions may help in the evaluation of a person's oral communicating ability, quality of grammar, attitudes, philosophy and many other tangible items that are essential requisites of a good teacher.

It is suggested that the third examination or manipulative skill performance consist of a listing of job or operations that are representative of the trade in which the examination is being given. From this list the examining committee may select one or more of these jobs or operations to be performed by the candidate. It would not be possible in a one-day examination to have the individual perform all the jobs or operations represented in the trade; however, the committee may select a sampling that would give them some standard of workmanship in measurement upon which to appraise or evaluate the person's skill in the operation.

Plan for Administering the Examination

There are many ways that the committee could administer the examination; however, experience has shown that one of the major problems is securing qualified personnel who can give sufficient time to help administer the examinations. For this reason it is suggested that a maximum amount of testing be done in a minimum length of time.

The following plan has been found to be very satisfactory. When the applicants arrive for the examinations, start them on the written examination. They may be called out of the written examination, in some designated order, for the oral and performance examinations. Since most schools require that a candidate must make at least "B" or better on proficiency examinations, the first item to be determined by the examination committee is, should the person be passed or not? The second item is, if they are to be passed, should they receive an "A", or "B" which may be considered satisfactory or excellent. In most cases if a person fails the examinations he is required to review and wait for a period of one year before reapplying for the examination.

Factors in Implementing the Plan

Even though it seems very simple to suggest a plan for the preparation, administration and implementation of a national plan of proficiency examinations for trade experience, there are many problems connected with the inauguration of such a program. One of the major problems is that of the preparation of comprehensive examinations that will substantiate the competencies needed for teaching technical grades and occupations. The simplest plan seems to be, since 38 states or colleges and universities are engaged in, or interested in, such a project, that participating members submit their quota of the examinations to a central pool which could duplicate and distribute them to all members.

For example, by contributing examinations for one trade, each participating state could receive 37 other trade examinations. To insure uniformity of quality and content of examinations, a format and set of criteria should be established for the preparation of these examinations.

The Problem of Granting College Credit

Even though we have moved into a technological age, there are still leading colleges and universities that are very traditional in practice. The idea of granting college-university credit for experience gained in a non-academic climate is very questionable in the minds of many academic educators. One of the questions that seems always to arise, in getting such a plan approved, is that of how you can justify giving college credit for experience that was gained off the college campus in a non-academic surrounding.

There are also those who view with alarm the possibility of losing national association accreditations for granting college credit for such non-academic accomplishments.

Another question which you may anticipate is how can you validate an examination and establish national norm without administering the examinations to thousands of teachers to establish these norms. There is an answer to this question. These examinations should first be validated by committees from industry who are thoroughly familiar with the requirements of the occupations represented in that industry. Once a comprehensive set of examinations in a trade or occupation has been industrially validated, each college or university could set up its own cut-off score as to passing or failing on the examination. National norms could be established later.

Another problem in inaugurating such a plan for giving college credit for industrial trade experience is that of securing the support of other departments and colleges throughout the university. For example, colleges of engineering have for a long time recognized the value of industrial experience. Their plan of cooperatively training students in industry attests to the value of industrial experience. Incidentally, the College of Engineering at the University of Tennessee lists 112 quarters that may be engaged in, by students, in industry. To date, they are very reluctant to consider any plan for, and will not give, college credit for this co-op training. Such a plan in industrial education might weaken their position. Other colleges or departments, such as Business

Administration which competes with private schools on business and secretarial training, will not support such a plan for giving college credit for experience gained off the campus. According to representatives from these departments, applicants could probably pass examinations on many of their courses, but it would be a dilution of standards. A department of industrial education seeking approval for such a plan is probably in a better position to secure support if it is in the College of Education and not in the College of Engineering or Business Administration. This arrangement would make it possible to secure approval for granting college credit in the College of Education only and not in other colleges or departments.

The usual procedure for securing approval on courses for granting this type of college credit is the preparation of a proposal which is usually submitted through the following channels. First it must be considered by the undergradaute committee. The success or failure in getting such a plan approved is largely dependent upon the composition of this committee. With all due respect to those who teach on the college level it is doubtful if you can explain to the complete satisfaction of all committee members the justification of such a plan. Since all phases of education, especially higher education, are very reluctant to pioneer any new practices in education, the best plan for getting such a program approved, is to show that other colleges and universities are already following such a plan.

After securing approval by the undergraduate committee, the next group to be considered is usually the faculty of the College of Education. If the plan receives the approval of this group, it then usually goes to the University Committee on Courses and Degrees. Since this committee is composed of representatives from all colleges throughout the university, it is here that much pre-explanation and work should be done since you may expect to encounter resistance from such schools as Liberal Arts and others who have little or no understanding of vocational education. Your strongest pleading at this point seems to be in pointing out the urgent need of vocational technical teachers, along with a list of other colleges and universities which are already participating in such a program. It has been the writer's experience that if you survive the first four hurdles, you are probably in the credit granting business.

The final approval of the plan must be secured from the University-wide Senate. This group is usually composed of all the deans of all the colleges. While it is not always true, the assumption at this level is that there has been much discussion on the matter to this point by others who have already given it careful consideration. There may be a need for further justification, however, approval by the Senate in most cases is a matter of formality. The department head may be called upon for more information; however, the dean of the school or college of education presents the plan to the Senate.

Place of Credit in a College Degree Plan

Since most four-year baccalaureate degree plans include 3 years of required courses and one year of elective courses, it is suggested that college credit gained through trade competency examination be given in the area of course electives. This arrangement seems to satisfy our academic friends, since it can be shown that teachers in the vocational curriculum are completing all the required courses that are required of other students in a similar baccalaureate degree plan.

Let us never forget that today children are being born into a world that is much different from the one in which their parents live. They will grow up in a world even different to the one in which they were born. They will mature and grow up in another world. They will die in a world very different to the one in which they live.

If education is as important as we think it is, the question arises, are we in education equal to the challenge of helping these individuals adjust, readjust and adjust again during their lifetime. It has been estimated that the average individual will need to be trained or retrained from 3 to 7 times during his mature life. In education as in industry, we must recognize that there may be a better way, or at least equally as good a way, of doing things than we have done them in the past.

Through experience we now have empirical evidence that the granting of college-university credit for industrial experience is an innovation that will help us do much in meeting the demands of change for more and better industrial teachers.

BIBLIOGRAPHY

BOOKS

Allen, Charles R., The Instructor - The Man and His Job, J. B. Lippincott Company, Copyright, 1919.

Hawkins, Layton S., Prosser, Charles A., Wright, John C., Development of Vocational Education, American Technical Society, Chicago, Illinois, Copyright, 1951.

PUBLICATIONS OF THE GOVERNMENT AND OTHER ORGANIZATIONS

U.S. Department of Health, Education, and Welfare. Administration of Vocational Education, Washington, D. C., U. S. Government Printing Office, 1958.

Earhart, Cecelia Ruth, The Kansas State Teachers College Bulletin, Pittsburg, Kansas, Ferd Voiland, Jr., State Printer, Topeka, Kansas, 1946.

UNPUBLISHED MATERIAL

Griess, Jerald A., Feasibility of Providing Trade Competency Examinations for Teachers on a National Basis. Rutgers - The State University, New Brunswick, New Jersey, 1966.

PERIODICALS

Keil, Melvin V. and Neubauer, John W., Definition of Terms, Phi Delta Kappan, April, 1965.

APPENDIX I

PREPARATION, ADMINISTRATION, AND IMPLEMENTATION OF TRADE COMPETENCY EXAMINATIONS FOR COLLEGE-UNIVERSITY CREDIT[1]

by

Ben S. Vineyard[2]

It is indeed an honor to appear before this distinguished group of educators to respond to the paper presented by Professor Joe Reed.

Professor Reed has done an excellent job in presenting his paper, and I am in general agreement with the proposal he has made for the "Preparation, Administration and Implementation of Trade Competency Examinations for College-University Credit." However I should appreciate an opportunity to elaborate on a few of the recommendations stated in the paper.

In the areas of implementation and administration of occupational competency testing, past experience indicates some additional suggestions are needed for persons planning such a program.

Unfortunately, before any plan of granting college credit for work experience can be implemented, a proposal must be legislated through a series of hurdles on the campus. Getting a proposal approved by the department curriculum committee, the college curriculum committee, and the university senate may pose problems similar to those experience by a member of the state legislature who introduces new legislation. Careful planning seems to be the "Key" to Success. It is most important to know the power structure in the committees and to solicit the support of the campus leaders at all levels. Informal techniques in getting your proposal approved may be far more effective than the strongest pleading that can be written. The time to present the proposal is also very important. I have observed that when a number of other departments are making curriculum proposals it is easier to get cooperation. A well planned proposal which is not in conflict with any university regulations has a much better chance of approval than one delayed at different levels due to small technicalities. A strong and cooperative State Board for Vocational Education can also be very influential in assisting teacher educators in the initiation and approval of programs for granting college credit for work experience. This is especially true if the State Board is willing to give financial assistance to the college or university.

A problem of implementation, once the approval has been granted, is establishing guide lines for operation. Probably the first question to be answered is, "Who is eligible to take the exams for college and/or certificate credit?" In my state it has been necessary to outline this information in the catalog and prepare instruction bulletins for distribution. Only persons who have the necessary work experience and other requirements to qualify for a certificate to teach trade and technical subjects in Kansas are eligible to take competency examinations.

---

[1] A Reaction to the paper of Professor Joe L. Reed.

[2] Dr. Vineyard is Chairman, Department of Trade and Technical Education, School of Technology at Kansas State College, Pittsburg, Kansas.

Many problems can be avoided if careful screening of applicants is made before approval to take the examinations is granted. Applicants should be carefully interviewed and required to present valid evidence of work experience. Records of apprenticeship, reports from employers and references are needed. Without adequate restriction, persons desiring college credit for one reason or another may make application to take the examination.

In the last few years I have received many applications which could not be approved. One chap majoring in another area requested permission to take the carpentry examination, hoping to receive twenty-four semester hours of A or B to raise his grade point average.

I agree with Professor Reed that high quality written, manipulative, and oral examinations are needed to provide a basis for granting college credit and/or certifying trade and technical teachers. The written test should probably be given first; persons successfully completing the written section should thus be given permission to take the manipulative examination. An oral exam dealing with technical information, in my opinion is not needed. However, an oral exam or interview is important in judging the applicants reactions to questions of a general nature and accessing his verbal ability.

Cost of Proficiency Examinations.

Establishing a fee for the competency examination may present a problem at some institutions. I am aware of a number of different methods used by various colleges and universities. Presently, we are charging the applicant no fee for taking the exams. The State Board for Vocational Education has provided money to pay persons giving the manipulative exams. The written exams are given once a year by the staff of the department of trade and technical education.

All written exams are of the objective type and provided with a key. The scoring is done by the office staff and recorded in the proper place.

I am not in agreement with Professor Reed's recommendation to include a large number of persons in the examination committee. In my opinion a large examining team or committee is to be avoided. We have used a committee composed of trade teachers to review and validate the tests; however, the administration of the test is the prerogative of the teacher-education faculty and person assigned to give the manipulative tests. The final results of the test may be reviewed by the State Supervisor, and the Dean.

It may be important to invite a number of persons to meet and interview each person who successfully completes the written and manipulative exam. This may be good public relations and increase the status of the testing program.

Types of Examinations.

I am hopeful that it will be possible to establish a more comprehensive testing program than Professor Reed has suggested. It would seem to me that with the interest and money available today it would be possible to develop written competency examinations in most of the trade and technical areas on a national basis. Large scale programs of testing are now conducted on a national basis for guidance programs and many other areas. I realize that developing proficiency tests in a number of trade areas will probably not be a profitable venture for large organizations such as Science Research Associates and the Educational Testing Service. However, with their professional help as directors in test building it may be possible for another agency to develop tests which will have the level of validity, reliability, objectivity, and discrimination needed for our work. Manipulative tests may be more difficult to develop, since the administration must be done by persons with varying levels of skill in this type of testing.

I also see many additional benefits for a program of proficiency testing. The granting of college credit for teachers may be one of the less significant uses of proficiency tests. In a few years achievement tests may be given to all students graduating from vocational programs. An organized national program should be able to provide services to both groups.

I am in full agreement with Professor Reed that a well organized and administrated testing program will improve the status of vocational teaching. Although the tests may not significantly predict teaching success, they will provide an instrument to evaluate the level of skill and knowledge possessed by our trade and technical teachers.

As you all know, a national testing program has negative aspects as well as positive. For years Federal and state civil service agencies have been giving tests for the selection of employees. Private agencies have found helping individuals pass civil service examinations to be a lucrative business. It is also possible to purchase printed materials which are advertised to assist students in passing college entrance examinations. In some European countries, passing exams is very important to the student and all types of exam scandals have developed.

If trade and technical competency exams should become extensively used, it must be expected that the tests will be used for purposes not intended. The tests would then need continuous changes with new forms developed each year. The expense of maintaining a testing program under these conditions would be increased considerably. Many of our trade and technical education teachers and supervisors would probably not be willing to accept a test score on an examination as the only basis for certification or granting college credit. I personally think final judgment should be made on the basis of scores on competency examinations and other factors including recommendations of the examining committee. Research on my campus has shown the predictive value of standardized tests given to be very low, and I fear the same situation could exist if competency examinations should be the only criteria for the granting of college credit or the certification of teachers.

APPENDIX J

THE PERFORMANCE PHASE OF TRADE COMPETENCY EXAMINATIONS

by

Benjamin Shimberg[1]

George Bernard Shaw's epigram, "Those who can, do; those who can't, teach.", is clearly a canard on the teaching profession, and especially for those teachers engaged in vocational education.

Long before vocational education found its way into the school curriculum, young men and women learned their craft by working alongside journeymen and master mechanics who were well versed in the skills of a trade or occupation. It was this tradition of learning from craftsmen that the founders of vocational education sought to preserve when they wrote into the original Smith-Hughes Act the stipulation that the instructors of vocational subjects be experienced craftsmen. Thus it was that experience, rather than formal education became the sine qua non by which vocational teachers qualified for their positions. To be sure, formal education requirements have been imposed in addition to the experience requirement, but experience continues to be the cornerstone of vocational education instruction.

Unfortunately the word "experience" -- like many words in our language -- has lost much of its precision because of careless usage. At one time one could assume with reasonable safety that the experienced craftsman was competent in all facets of his craft. Today, it is no longer safe to operate on that assumption. The concepts of division of labor and specialization of function have altered the structure of industrial society -- and the nature of the people who work in that society.

Men who may have had a well-rounded repertoire of trade skills when they completed training have often seen many of their skills wither and all but disappear as they have honed certain other skills to a fine edge of perfection. The all-around auto mechanic becomes a transmission specialist; the all-around printer does page makeup or sets display ads. Years of experience may win him seniority and higher pay, but these have little to do with his all-around trade proficiency.

This poses a serious problem for vocational education. Administrators have stipulated that vocational courses must be taught by men with experience, but they also would like these men with experience to be proficient in all aspects of the trade or occupation.

Unfortunately, it has been easier to obtain information about years of experience than about "trade competency" so that the former has gradually edged out the latter as the basis for determining one's eligibility for certification.

This is not to say that the importance of assessing proficiency has not been recognized. An unpublished survey by Schaefer in 1959 revealed that 16 states were making use of proficiency examinations to some extent. A more recent study by Kanzanas and Kieft at Eastern Michigan University elicited information about trade proficiency testing in 11 states.

---

[1] Dr. Shimberg is Director, Vocational-Technical Education Projects, Educational Testing Service at Princeton, New Jersey.

At the first session of this seminar Lofgren reported on the extensive use of trade competency examinations in California. Reports were also heard about the testing activity going on in New York State, Florida, and Michigan.

In each of these states, as well as elsewhere, written examinations are being used to assess the cognitive aspects of a trade: the job knowledge, theory, and related mathematics and science which provide the essential background for effective practice as well as for successful teaching. But there is also recognition of the difference between knowing about a trade or occupation and being able to perform the major functions of that occupation at a satisfactory level of competence. It is for this reason that tests of occupational proficiency -- in addition to written tests of knowledge -- are assuming increased importance in the area of occupational assessment.

For many years the belief was widely held that there was such a high relationship between trade information and performance that the former could serve as an indirect measure of the latter. This idea may have gained currency in an era when research showed a generally high correlation between written tests and final course grades which were uncritically accepted as a valid indicator of overall proficiency. Since course grades were themselves often based on written tests, it is not surprising that this relationship obtained. However, when special attention was devoted to assessing shop performance and such performance was given appropriate weighting in the final grade, the relationship tended to go down substantially. Writing about the Navy's experience with performance measures during World War II, Stuitt says: "Although it had been assumed that written tests sufficed to indicate what a man had learned in a service school, the evidence showed that performance tests and improved shop grades were not closely correlated with written test grades. During tryout in a gunners' mate school, performance tests correlated from .14 to .35 with written tests and only slightly higher with final grades which were based largely on written tests. In a torpedoman school where shop grading was quite good, test tryouts showed that, on the average, their sample performance tests correlated .63 with final grades, but only .38 with the multiple choice final examination. (p. 306)

Today it is generally conceded that written tests of trade knowledge are not a very dependable way to evaluate shop performance and that without some type of direct or indirect performance measure it is unlikely that we can make an accurate assessment of an individual's trade competency.

The most commonly used approach to assessing performance is through a work sample. This requires that the individual being tested demonstrate his knowledge or skill by completing a series of tasks or a segment of work under actual conditions in the work situation. It may properly be thought of as a controlled tryout under actual work conditions. A mechanic may be told to diagnose and repair a malfunction in an auto. A student of TV repair may perform a similar task on a set into which a number of "bugs" have been built. Such tests come about as close to duplicating the real life situation as possible. In a sense they are criterion tests since they involve the desired behavior almost in its totality. Since it is frequently impractical or uneconomical to require the performance of a complete sequence of behavior, a more limited sample may be selected so as to be predictive of the behavior as a whole.

"Simulation" is a general term applied to a training or testing situation which imitates the actual work situation in a realistic way. Fraser, writing about the use of simulation in aerospace training suggests that "simulation is the art and science of representing the essential elements of a system out of their usual setting in such a manner that the representation is a valid analogy of the system under study." (p. 2) As applied to performance testing, simulation tests seek to isolate and duplicate essential features of a task or operation. To the extent that they succeed in capturing the essence of the criterion task, they have many advantages over work sample tests: ease of administration, economy, convenience, and safety. However, some simulated situations have face validity (look good) but fail to measure the essential characteristics of the job. Thus, all efforts to utilize simulation as a testing device must be carefully validated against the criterion performance.

Whether one undertakes to assess performance with a work sample test or by some type of simulation, the first step, in either case, is determining what to test. Great care must be exercised to insure that the tasks assigned are representative of the desired (criterion) performance. The choice of tasks should be made in the light of a thorough knowledge of the job as a whole. This will usually involve a job analysis for the identification of the critical skills inherent in the job. It is not uncommon for individuals involved in developing performance tests to undergo actual training in the skills area in order to obtain first-hand experience with the requirements of the job under study.

While the job analysis will reveal a great deal about the scope of the activity, the nature of the tasks performed, and their relative importance, we know that in general it will be possible to include only a limited number of tasks on the performance test. Unlike the written test, however, (where one may sample widely the knowledge, skill, and understandings of an individual) the performance test developer must rely on a very small sample of behavior. This places a high premium on selecting a sample of tasks that is representative of the job as a whole. His choice of tasks must be guided not only by an awareness of what is truly important and critical, but also what is practical and feasible in terms of time available, equipment, cost of materials, and personnel.

Let's look at how one group of test developers approached the task a number of years ago. Psychologists at the Institute for Research in Human Relations were awarded a contract by the Office of Naval Research to explore the development of practical performance measures. One of the jobs for which they undertook test development was that of the Aviation Structural Mechanic. Initially, they analyzed training manuals to ascertain job requirements (Presumably, these had been based on a careful job analysis). Discussions were then held with a number of Chief Structural Mechanics to identify jobs that might be used to measure the various requirements. In all, some 66 tasks were suggested. These were put on 3 x 5 cards and rated by the Chiefs. They were instructed to identify and to eliminate certain tasks; those that would take more than an hour to complete; those that required material or equipment not usually found in operating squadrons; those that would be costly in terms of material; those that might cause interruption in naval operations; and those that involved a great deal of repetitive activity, such as removing and replacing a whole series of nuts and bolts. Consideration was then given to the problem of measurement. Those tasks that did not seem amenable to objective evaluation were eliminated.

To obtain some indication of "validity", the Chiefs were asked with respect to each task, "Would you be willing to assign a man to this task (eg: welding) after seeing him complete this (welding) job?" Only those tasks to which all the Chiefs answered the question affirmatively were used in the final battery.

Ryan and Fredericksen mention a number of these as well as other considerations in their discussion of performance tests. They suggest:

1. The sampling of activities should be as wide as practical.

2. A minimum of easy or routine operations should be included.

3. The task should be sufficiently exact to permit accurate standardization and enable objective judgments to be made.

4. The task chosen should have face validity to command the respect of the examinee.

5. Tools and equipment should be reduced to a minimum and should be capable of standardization.

While considerations such as these impose practical limitations on the tasks that may be assigned, within these limits the final determination should be made in terms of how well the tasks represent the job as a whole. Every effort should be made to maintain a proper balance among the various elements revealed by the job analysis. In the case of the Aviation Structural Mechanic Job, three elements were considered critical: repairing, replacing, and trouble-shooting. In selecting a manageable group of tasks, the test developers sought to tap as many as possible of the diverse skills that this specialist is called upon to use. One task required the examinee to fabricate a flush patch for stressed metal. He had to demonstrate not only his skill in metal work, but also his ability to read blueprints, to use measuring instruments, and to do riveting.

A two-way grid, listing job specifications along one axis and performance tasks along the other will help to identify the contribution of each task to the evaluation process. Such a grid will serve to reveal gaps, overlap, and possible duplication. While it may not always be possible to complete the grid, it will focus attention on important elements that might otherwise be overlooked.

Once a number of performance tasks has been identified, the decision remains to be made how performance is to be evaluated. In attempting to arrive at such a decision with respect to the Structural Mechanic Battery, one Chief said, "I don't care if a man stands on his head while doing a job, as long as it's OK when he's finished." This Chief might be described as "product-oriented". He was concerned only with the precision of the final product and with its correct operation. While most evaluators would agree that "quality of the final product" is of great importance, they are likely to argue that some consideration should be given to the "process" by which the final product is obtained. They would evaluate the individual's care of equipment, his observance of safety rules, and his adherence to approved methods. They might also take account of the amount of material he wastes and the time he takes to do the job.

In practice the relative weights to be given to "process" factors and to the "end product" will depend on the objectives of the test and the nature of the task. Evaluating "process" is a time-consuming and expensive procedure. Great care must be exercised in developing the rating forms and in training observers. Even so, results may not be as dependable as we would like because of subjective factors beyond the evaluator's control. Questions naturally arise as to how much importance should be attached to "process" ratings. In the original planning for the Structural Mechanics' test, equal weight had been given to "process observations" and to "final product ratings". The Chiefs objected. They pointed out that a man might do all the right things ("process") yet wind up with an unusable product. They insisted that substantially greater weight be assigned to "product ratings" than to the "process ratings".

In all likelihood this battle must be fought anew each time a performance test is developed, for it involves a value judgment that can only be made by those responsible for program design. There is justifiable concern, for example, that correct procedures and safety considerations - stressed in the instructional program - will be undermined if "process" is ignored by the evaluators. Thus, if evaluation is perceived as an integral part of instruction, then this viewpoint is certainly defensible. If, on the other hand, the purpose of evaluation is to predict subsequent on-the-job performance, the major consideration should be how much the "process" score contributes to the overall validity of the performance test. Unless higher validity can be demonstrated, it is questionable that the effort and expense can be justified.

The actual design of performance tasks generally calls for ingenuity and imagination. This is less apt to be the case when work sample tests are used. However, even here the evaluator is faced with the necessity of conserving time; hence he will seek out tasks which call for the display of critical skills and which minimize the routine aspects of a job. The examinee may be given a partially finished piece of work and instructed to complete the job according to blueprint specifications. If he can perform the exacting jobs associated with finishing operations, it may be assumed that he could have done the preparatory work had he been asked to do so. In a similar vein, it would make sense to eliminate activities involved in getting to the critical job. These may consume time, yet involve little more than the removal of numerous screws, nuts and bolts. Where the total job might take a long period for completion, it may be broken down into sub-tasks which can be performed in a reasonable length of time.

As one gets away from the "live" work situation and moves into simulation, the possibilities for applying creative imagination increase. One line of development is the construction of equipment which has the essential operating features of the "real thing", but is far less complex and therefore less expensive. Almost at the other extreme are the high fidelity simulators used in training jet pilots and astronauts. Here the simulator is generally very complex and very expensive because of the great premium placed on reproducing as faithfully as possible in the training situation as many as possible of the conditions one is likely to encounter in flight or on a space mission. We shall not concern ourselves at this time with the application of high fidelity simulation in the area of trade competency examinations. In time, offshoots from the high fidelity field are likely to trickle down to the more mundane level of skilled trade evaluation. However, it hardly seems profitable to speculate when or how or in what ways this may occur. However, it does seem reasonable to assume that computers will be put to use as a technique for checking out trouble-shooting or problem solving ability. There have, in the past, been numerous attempts to devise tests which describe a malfunction in a piece of equipment and ask the examinee what he would do to track down and correct the condition. For each course of action selected, the examinee gets feedback -- information as to the outcome. On the basis of this information, he selects the next step -- and so on, until he has solved the problem. This approach has been used in the form of a "tab test" (one lifts a "tab" to discover the outcome of an action). It could also be presented in scrambled book form. The computer seems ideal for such tasks since it could be programmed for an almost infinite number of possibilities, and it could keep track of the entire sequence of operations made by an examinee.

While computers and other types of sophisticated hardware may one day serve in the assessment of trade performance, it would have to be shown that the skills a person had demonstrated on the computerized test were actually related to success in a given job. Even in trouble-shooting, there would seem to be a difference between knowing what to look for and actually performing the tests to track down and correct a malfunction.

The Institute for Research in Human Relations also conducted an investigation of the trouble-shooting ability of Aviation Electricians under contract with the Office of Naval Research. One of the tests they developed involved the Aviation Electrician's ability to perform a series of electrical checks using the multimeter as a test instrument. A testing box was constructed to simulate the type of electrical circuits found in operational aircraft. Letters (A-O) were used to indicate terminals at which readings were to be taken by examinees, and numbers (1-8) were used to denote the resistors. This 60-item test had a split half reliability of .73 which increased to .84 when the Spearman-Brown correction was applied.

Another simulation test developed for this project was called the Basic Skills Test box. The box simulates, in miniature, an aircraft section and contains a simulated motor, control relay, control cable, ribs, fuel line, junction box, and lightening holes. The task of the examinee was to solder wires to the 8-pin cannon plug and to run wires through their components as indicated on a schematic. The abilities involved in this task were: Soldering, use of tools, knowledge of principles of safe wiring, selection of proper size nuts, bolts, and clamps; reading and working from a schematic, etc. The reliability of this test was determined by correlating two rationally equivalent halves (using criteria suggested by Thorndike.) For a group of 15 students the split half reliability was .56, which was raised to .72 when the Spearman-Brown correction was applied.

Many other examples could be cited of "black box" simulators which required the examinee to demonstrate that he could apply the skills he had learned to realistic problem solving situations. More frequently, the performance task does not require a black box. To find out if the Structural Machinist could fabricate a flush patch in the fuselage of an airplane, he was given a piece of aluminum with a slight hole in it and was told to repair the damage. The examinee had to calculate from a schematic diagram how far apart to set rivets and how close they should be to the edge of the metal. To conserve time, he was required to do only half of the riveting.

A test of the examinee's ability to fabricate a rigid tubing assembly involved the use of a "mock up" with a standard fitting at the back and another fitting at the bottom. His job was to fit the tube to the standard fittings. To do this he had to cut, bend, flare, and fit an 18" length of specified pipe. The bends were typical of those found in aircraft. Since the "mock up" box was closed on three sides, the examinee was forced to work in relatively confined space, such as is usually found in an aircraft.

These illustrations have not been selected as models of how performance testing should be done. Rather, they have been brought in to emphasize the distinction between actually *performing* a job -- even under simulated conditions -- and *responding to questions about a job*.

There is almost no limit to the possibilities for developing performance measures. In some cases training equipment may be adapted to test purposes. Circuit boards used in teaching electricity would seem to lend themselves to this purpose. The motor analyzers found in most auto repair shops could probably be "programmed" to simulate a variety of automotive malfunctions. New approaches may also be devised to get at critical skills which are developed in a training program, but which are seldom tested systematically. For example, Dr. Thomas Baldwin at North Carolina State, is developing a series of auditory tests for auto mechanics. Certain malfunctions will be built into an automobile and recorded on stereo equipment. Presumably, master mechanics will be able to identify the nature of the malfunction from a sound recording more readily than mechanics in training. Both of these groups should earn higher scores than students who have not had training in the field. Dr. Baldwin is also planning to develop tests of certain kinesthetic abilities which he believes are developed in trade training programs.

Whatever the nature of the task may be, sooner or later the problem of evaluating the performance must be faced. We mentioned earlier that one may wish to focus exclusively on the quality of the product, on the process, or on both the product and the process.

"Product evaluation" is easier to deal with than is "process evaluation". For one thing, the product is often a tangible object, more durable than the fleeting actions which make up a process. Such a product may be judged after the testing has been completed. Process evaluation on the other hand must generally be done while the testing is in progress.

In the case of a tangible product it is generally easier to obtain more reliable judgments regarding quality than one can for a process. If the product is one that has been made to precise specification (such as those found in a blueprint) it is possible to check how closely the product conforms to the specifications. However, one should not overestimate the ability of judges to evaluate such a product, even when precise specifications are available and the judges use fine measuring instruments to check for accuracy. During World War II, four instructors were asked to assess the quality of 30 "samplers" prepared by students in a basic machinist course. Although the judges used appropriate instruments to make their evaluations, there were many discrepancies among the grades assigned to the same samplers. Judges' rating intercorrelated from .11 to .55. Then a set of taper gauges and caliper gauges was devised with scales for five points of deviation on either side of specifications. When these gauges were used in scoring the samplers, the ratings correlated .93 on the one set of samplers and .96 on another (p. 306). This suggests that every effort should be made to make product evaluation as objective as possible, and that jigs and gauges may be useful for this purpose.

In situations where quality must be judged subjectively, it is important to list the characteristics which differentiate the good from the poor product and to devise techniques for measuring or otherwise assessing these characteristics. It is sometimes possible to increase the reliability of judgments by developing a comparative scale. This may be done by having a group of highly competent judges place a number of "products" in rank order on the characteristic being rated. When a stable scale has been created (to provide benchmarks for differing degrees of goodness) it may then be used by less qualified judges to ascertain where along the scale a given product fits.

When the end product is a service -- such as the repair of an auto or a TV set -- the judgment is generally in terms of utility. Does it work? and how well? However, it is unlikely that we would be satisfied to know merely that an examinee effected the repair. Part of our evaluation would hinge on how long it took, and whether the solution was the most efficient one for the situation. Such questions inevitably take us over into the area of process, for we are now concerned with how the job was done, not merely with the end result.

As we indicated earlier, how much importance to attach to process evaluation is a value judgment which depends in large measure on the purpose of the evaluation. In selecting a machinist who can meet exceedingly precise specifications, one might be less concerned with his procedures or with the time required than with the end result. However, in selecting an instructor for a vocational program, one might attach considerable weight to "process" as well. In each instance the importance of "process evaluation" must be weighted against the effort and expense involved in obtaining such information.

When "process" information is deemed to be an essential part of evaluation, great care should be exercised in defining specifically what type of information is needed. The specifications for process evaluation should emerge from the job analysis. How important is speed? Accuracy? Use of approved methods? Care of tools and equipment? Adherence to safety standards? There is no point in burdening the observer with assessment of procedures which are not of critical importance. By focusing only on the essential elements of "process", the chance of getting reliable judgments is increased. Requiring extraneous observations is almost certain to be at the expense of overall accuracy.

After the process dimension has been defined, a rating form should be devised. The form is essentially a check list covering each step of the process and providing criteria for making process judgments. Stuitt reports that in devising rating forms for performance in various navy enlisted schools "sheets which allowed considerable leeway in evaluating quality of performance were found, in general, to be unreliable because different instructors did not agree in grading trainee performance. For objective scoring the proctors' check sheets were made highly specific" . . . (p. 300)

The problem of rater reliability to which Stuitt refers has continued to be a major drawback in "process evaluation". Unless reasonably uniform rating standards are adhered to by all observers, it is impossible to disentangle the variance in scores attributable to differences in examinee performance, from differences due to rater "performance". Training of observers deserves much greater attention than it has received in the past. Even then, it is questionable how much uniformity can be achieved. The experience of the College Board in training teachers to grade CEEB essays suggests that only small gains in reliability may be expected even when a substantial effort has been made to train the raters.

The advent of video-tape recorders may offer a fruitful approach to training of observers. After some training in the use of the observer's check list, prospective observers could be shown a video-tape recording of an examinee taking the test. After these observers had completed their ratings, discrepancies among raters could be discussed and differences resolved. Then the group could be asked to rate a second set of video-tape recordings. Those observers who persisted in making divergent ratings might be exposed to further training or dropped from the roster of qualified observers.

In some situations where "process evaluation" is of great importance it may be possible to record the entire test performance on video-tape so that two or more observers could rate the individual after the performance had been completed. Where ratings differed, judges could re-examine the tape together, discuss the behavior in question, and resolve their differences; or a neutral judge could be called in to assist in reaching a decision.

In this paper there is no need to dwell on the obvious need for developing detailed instructions for both administrators and examinees, for carefully defining the work situation, and specifying necessary tools and supplies, etc. It should also be obvious that the directions, performance measures, and rating procedures must undergo thorough pretesting followed by careful analysis and revision. Part of the analysis should include a check on validity. This poses serious problems, since a carefully developed performance test is likely to be a more valid criterion measure than the usual criteria, such as supervisors' ratings or years of experience in an occupation. Nevertheless, we need some assurance (beyond face validity or expert judgment) that the test does, or can, in fact, differentiate the journeyman from the apprentice; the recognized specialist from the marginal worker. If we find unexplained reversals in the performance of carefully selected criterion groups, we may wish to reexamine our tests or rating procedures. We may be concentrating too much attention on fine detail which the proficient worker does not readily recall or which he is seldom called on to use. We may be missing the essential skills that differentiate the highly proficient worker from one who is merely satisfactory.

We should have no illusions about the difficulty of carrying out studies on the reliability and validity of performance tests. The problems of performance evaluation are infinitely more complex than those encountered in written tests. The skills of highly competent measurement specialists working closely with experts from the subject field are needed to devise new approaches that will insure that these tests approach professional standards.

Much of the work that has been done in the evaluation of occupational proficiency (outside of the armed services) has been severely hampered by lack of adequate resources. Funds have not been available to employ professionally qualified personnel to work on the construction of the instruments, or even more important - on the analysis of these instruments to insure their reliability and validity.

When one considers the importance of performance evaluation to the future of vocational education it seems inconceivable that so little progress has been made. The pioneering works of Lofgren, Hankin and others stand as tributes to what dedicated men can accomplish with minimal resources. However, this does not argue that we should try to get along in the future, as we have in the past, on a shoestring budget. Progress depends on going beyond what these pioneers have been able to accomplish during the lean years before vocational education was catapulted into national prominence. Resources must be found to bring to bear on the whole process of performance test development the highest levels of psychometric skill and theory that we are capable of mustering. Such a concentration of skills is not likely to occur as long as various states continue to work independently on the problem. A coordinated research and development effort - in which many states pool their resources - would help to eliminate wasteful duplication and provide support for exploration of creative new approaches.

The idea of nationally-administered proficiency exams poses many problems not present when testing is done within a single school or plant. Variations in equipment from one test center to another is only one of the difficulties that must be overcome. It may turn out that the only feasible approach would be to have a limited number of centers and to require all applicants to travel to the centers for evaluation. On the other hand, we may be able to devise simulation devices which can be standardized and which will yield reliable measures of proficiency. It would, of course, be essential that such devices be validated against criteria of actual performance. Thus, highly reliable performance tests might serve as criterion measures, while simulation tests would be used as predictors. It would seem that such avenues need to be explored very carefully as we enter a new era in the history of vocational education and the measurement of occupational proficiency.

BIBLIOGRAPHY

Conrad, H. S. Summary Report on Research and Development of the Navy's Aptitude Testing Program. Office of Scientific Research and Development, Princeton, New Jersey, 1945. 69 pp. (Restricted)

Fraser, T. M. Philosophy of Simulation in a Man-Machine Space Mission System. National Aeronautics and Space Administration, Washington, D. C., 1966. 107 pp.

Human Factors Research, Inc. Research on the Development of Shipboard Performance Measures and Performance Judgments. Santa Barbara, California, 1965. 21 pp.

Kanzanas, H. C. and Kieft, L. D. An Experimental Project to Determine More Effective Teaching Certification Procedures in Michigan by Competency Exams. Eastern Michigan University, 1966.

Ryans, David G. and Fredericksen, Norman. "Performance Tests of Educational Achievement" in Educational Measurement, E. E. Lindquist, editor. American Council on Education, Washington, D.C., 1951. pp. 455-494.

Siegel, Arthur and Courtney, Douglas. Development of Practical Performance Measures. (under contract with Office of Naval Research). Institute for Research in Human Relations, Philadelphia, Pennsylvania, 1953. 225 pp.

Siegel, Arthur; Jensen, John J., and Danzig, Elliot R. An Investigation and Test of Trouble-Shooting Ability of Aviation Technicians (under contract with Office of Naval Research), Institute for Research in Human Relations, Philadelphia, Pennsylvania, 1965. 80 pp.

Stuitt, Dewey B. (Ed.) Personnel Research and Test Development in the (U.S.) Bureau of Naval Personnel. Princeton University Press, Princeton, New Jersey, 1947. 513 pp.

U.S. Army Signal School. Performance Test Construction. Fort Monmouth, New Jersey, 1958. 33 pp. (Restricted).

APPENDIX K

THE PERFORMANCE PHASE OF TRADE COMPETENCY EXAMINATIONS[1]

by

Paul V. W. Lofgren[2]

There is an old axiom to the effect that no one can take a psychologist's theories apart any better, or more gleefully, than another psychologist. This seems to be the favored indoor sport in our fraternity, as well as the basis for whatever advancement made in the profession.

In reading Dr. Shimberg's report it occurred to me that we must have read the same literature and have drawn somewhat similar conclusions regarding the theories underlying the major types of proficiency tests he has reviewed.

May I take this opportunity to compliment Dr. Shimberg on his very concise but still meaningful treatment of the subject. Because of the prescribed briefness of the report much has obviously had to be left unsaid but to say as much as he has in so few words is, indeed, an art in itself. Also, I wish to publicly acknowledge Dr. Shimberg's kind mention of my name in his report as one of the pioneers in the proficiency testing movement in the vocational teacher selection field.

For some reason, most likely because the postal service gets overloaded this time of the year, the report arrived at my office on Monday of this week. Consequently, since some corners had to be cut in order for me to depart from San Francisco yesterday morning, there is only one extra copy of my reaction report available at the moment, Mr. Chairman. Also, you will find no annotated bibliography nor the meticulous organization that characterizes Dr. Shimberg's paper. Some of these shortcomings can, of course, be remedied with a few additional hours at my disposal.

Taking advantage of my new title of "pioneer" my reactive contribution, if such it can be called, will be in the form of amplification and exemplification of the pros and cons enumerated in Dr. Shimberg's report; and this largely in terms of personal research experience in the theoretical as well as the applied phase of "simulation" testing.

Taking the topics of the report in sequence, so far as possible, my reactions are as follows: At the beginning of his report Dr. Shimberg identifies a problem of semantics, namely, that the word "experience" has lost much of its precision because of careless usage. This is certainly true and reminded me of how the term "IQ" suffered a similar fate after World War I and became, for a time at least, almost meaningless. A few years ago a California billboard agency proudly advertized a beer with a high IQ (It Quenches).

---

[1] A reaction to the paper of Dr. Benjamin Shimberg.

[2] Dr. Lofgren is Supervisor of Occupational Proficiency Testing, Division of Vocational Education, Trade-Technical Teacher Education at University of California, Berkeley.

However, it occurs to me that states lacking a selective testing program and relying entirely upon employer recommendation and duration of union membership are the ones affected by this situation to a far greater degree than states equipped to measure occupational competency. The fact of specialization within an occupation has never posed a problem in California beyond that of increasing the workload on the testing division. Fortunately, California issues a Standard Designated Subjects Credential (SDS) specifying the exact subject area and/or subject limitation involved, viz. "Radio Communications (Ltd. Telephone only)"; "Dental Assisting (exclusive of Laboratory)"; "General Printing (Exclusive of Linotype)", etc. This prevents misrepresentation very effectively.

On the question of written vs. manipulative tests (p. 2, para. 1 & 2) a proposal was made in California about 10 years ago to abolish the manipulative phase of the test, as an economy measure, and to rely entirely on the written test. Since it could be shown that the intercorrelation never exceeded 0.30 the proposal was withdrawn. Dr. Shimberg's mention of spurious correlations giving rise to such proposals is most timely. This occurrence is all too often overlooked by the statistically unsophisticated.

In passing I wish to comment on one statement in paragraph 2, namely that: "The student of T.V. repair --- etc." Realizing that several states do, or intend to, apply occupational proficiency test scores toward a college degree I wish to repeat my answer to a question posed at our last meeting. Some of the California state colleges accept for credit certain practical experiences as evaluated and recommended by a State committee. Proficiency in subject matter, demonstrated by test grades, is rewarded by a certain number of points per grade obtained. However, since one of the requirements for evaluation is a minimum of 1600 hours of successful teaching (for which credit is also given) the "student" aspect never enters into our test deliberations, nor does teaching ability, but subject matter competency alone. The candidate must hold a high school diploma and be able to verify 7 years of occupational experience before being considered at all. The average experience level "across the board" is presently 13 years. Teaching ability for which there exists no test as such, is assessed partly by a conventional composite test and partly by observation of personality and performance which attending our teaching techniques courses. As you may know our SDS credential is issued with certain requirements on a deferred basis, including the required junior college Associate of Arts degree. At present approximately 50% of our candidates hold an AA or higher degree. The teacher receives a "clear" (life) credential only upon removal of all deficiencies. In the meantime he is employed by the school on a probationary basis.

The work-sample test (p. 2, last para.). The description includes, in one short paragraph practically all the reasons why I am very much biased in its favor. The statements of particular significance are: "This (test) requires that the individual being tested demonstrate his knowledge and skill by completing a series of tasks or a segment of work under actual conditions in the work situation." "It may properly be thought of as a controlled tryout under actual work conditions." "Such tests come about as close to duplicating the real life situation as possible," and, "In a sense they are criterion tests." These statements sum up most adequately the reason for my insistence upon work sample tests in California.

I have found "face validity" to be of prime psychological importance to the candidates, reflected in their approach to the test assignments as well as in their post-facto comments. Furthermore, if time and budget should ever again become a problem in vocational education there would be no interruption in the testing service because test equipment is available in our vocational school programs.

Limited sampling is, of course, inherent in all simulation tests. One must always be conscious of the fact that there is no absolute substitute for a long term "test" such as six months or more, of probationary employment. The only excuse I have ever discovered for using tests at all is to save somebody time and money. So we attempt to estimate competency by means of spot checking. The isolation of the appropriate "spots" to be checked is, in my estimation, on par with rating reliability in its importance. Dr. Shimberg has stressed this phase of testing in his report and I fully concur. In this connection I have found the Viteles psychograph technique helpful since it embodies the concepts of "brain vs. brawn" and percentage of total work time devoted to a given operation. The psychograph is closely allied to, and should be a part of, the job analysis mentioned on p. 3 of the report, para. 2, 3, 4 and the two-way grid described on p. 4, para. 5. These combined techniques, with the aid of competent occupational advisors, tend to isolate salient points to be tested, i.e. test activities that permit the examiners (judges) to infer from what they observe that "if the candidate can do this we can afford to assume that he also has the knowledge and skill to perform the antecedent operations as well as those that ordinarily follow." One example is the graining of a panel in the painting test. It is reasonable to believe that a candidate doing a creditable job on this would also know how to apply the undercoats and the transparent finish. Therefore, test time may be saved for an additional spot check job.

In many instances, though not in all, it is possible to anticipate the final result without completing the entire sequence of operations, especially when the "reward" value of completion does not in itself constitute a psychological problem to the candidate. For example, in Auto Mechanics where one of the job assignments may be to set up for reboring a cylinder. The judges observe the candidate's application of appropriate micrometers and note the readings he obtains; the boring rig is mounted and adjusted; finally the candidate turns the switch to activate the grinder. After a few turns, perhaps 1/8", the judges may order the operation "cut" because they know from what they have already observed what the end result will be.

With reference to another "simulation" test, the analogous, (p. 3, para. 1 & 2) may I mention that I once upon a time wrote a doctoral dissertation under the title "The Analogous Aptitude Test in Theory and Practice." For this task I accumulated a reference bibliography of 420 pertinent titles of books and periodicals, in English as well as foreign languages, going back to the first readily available publication on the analogous concept in testing by Hugo Münsterberg in 1913.

I am as convinced now as I was then that there exists no comparable test instrument for the prognosis of potential ability with reference to manipulative skills. I am equally convinced that the analogous test in most instances is decidedly impractical for the type of testing we are discussing here, namely, testing of achievement, or ability after training has taken place. Now, not to appear hide-bound beyond redemption I only wish to say that the evidence brought to my attention so far has been insufficient to convince me of the practicality of analogous tests for general application to the problem facing us. One of the major obstacles is the time-and-cost factor. I spent 2000 hours in gathering test and criterion data alone for my dissertation using arc welding as the prototype test.

Dr. Shimberg states (p. 3, para. 2) that: "It is not uncommon for individuals involved in developing (such tests) to undergo training in the skills area in order to obtain first hand information ---". To this I would add that such experience is well nigh imperative in order to develop a highly efficient analogous test. I learned to arc weld in three positions in order to observe the rate and sequence of the elimination of irrelevant movements during the learning process. In preparing to develop an analogous test battery in plumbing I attended a short term training course to become familiar with practically every manipulative phase of the trade. This included such significant aspects as that of caulking the hidden back side of a 6" pipe. In this activity the kinesthetic sense alone can determine the quality of the job. As my final example for today, I drove a streetcar up and down Mission Street in San Francisco for a week to determine what cognitive and motor response patterns must be coordinated in order to safely operate the power and brake controls of a streetcar while at the same time stomping on the warning bell button, pulling the signal cord, watching out for automatic switch junctions, stray dogs, cats, boys on bicycles, and old ladies crossing the street. All of this was for the purpose of developing appropriate aptitude tests, not tests of acquired ability.

I once had a sad experience with an analogous automobile driving test that was used as an achievement test by one of our metropolitan police departments.

Admittedly I am not too familiar with the use of the analogous achievement tests used by the armed services. On the other hand, what I have observed, heard, and read about their analogous aptitude tests and mock-up training equipment has impressed me greatly.

According to Wm. Koeler there are three major types of "synthetic" tests: the work-sample, the analogous, and the miniature. The word "synthetic" was coined by Koeler in his book Gestalt Psychology, 1929. Presumably, on my part, the terms "synthetic" and "simulation" carry the same connotation. The concept is, of course, not new. In this connection there is some doubt in the minds of many scholars as to whether there exists a single new idea in the world today. Every idea, or concept, that comes to one's attention seems to have a history which can quite readily be traced in a well stocked library. I have personally traced the idea of "synthetic" tests only so far as to the beginning of this centruy when the "father of applied psychology", Hugo Münsterberg, reported his findings. Where he acquired the idea I do not know at the present time but most certainly it was not original with him.

While the miniature "synthetic" or "simulation" test is not mentioned in Dr. Shimberg's report it was mentioned as a possibility for our purpose by one of the speakers at our first seminar. For this reason I take the liberty of casting a negative vote; this time on predominantly psychologic rather than economic grounds.

Frank Watts, "Journal of Applied Psychology" 1921, cited by Moore and Hartmann in Readings in Industrial Psychology, 1931, had this to say: ". . . as far as physical labor is concerned, there is reason to believe that ability to perform the fine movements called for in working with small models is not at all indicative of ability to perform the larger movements of the actual work which the small model is intended to represent. In the two cases not only will different muscular and nervous coordinations be necessary but also different types of interest. Thus the watchmaker and the miniature painter would usually be completely unsuited temperamentally --- as well as physically for employment respectively upon steam turbines and motor generators, or upon big poster work.

Preceding Frank Watts, Münsterberg, in his Psychology and Industrial Efficiency, 1913, took his stand as follows: "A reduced copy of an external apparatus may arouse ideas, feelings, and volitions which have little in common with the processes of actual life --- On the whole I feel inclined to say from my experience so far that experiments with small models of the actual industrial mechanism are hardly appropriate for investigation in the field of economic psychology."

My own experimental work with miniature tests has convinced me that the judgement of the two old timers I have just quoted is still sound. What puzzles me is that with the wealth of literature we possess people persist in repeating identical mistakes. I am not speaking against experimentation for its own sake but against untenable promises held out to an unsuspecting public.

In 1951 I was invited by one of our larger states to make a survey and evaluation of the facilities, equipment, and procedures employed by its State Board of Plumbing Examiners. Their facilities and equipment were excellent beyond compare. Apparently money was no problem. A standard procedure of test administration had been established and appeared to conform to professionally acceptable practices. The test battery consisted of a written test and a pictorial; bench work involving cutting, threading, and reaming lengths of pipe to exact measure, caulking, lead bending, lead wiping, and the adjustment of a gas burner flame. So far so good. However, I found that approximately 1/3 of the floor space was taken up by seven miniature two-story houses representing completely roughed-in wood frame construction. The roof and the two stories were detachable. The miniature structures were made to scale and, together with miniature aluminum fitting, pipes, and wooden dowels, constituted a trade knowledge test of sanitary plumbing. The candidates were furnished a special reduced scale ruler 1:5. This was one of the finest miniature tests I have seen. Unfortunately, my recommendation regarding this phase of the test, in which the examiners took great pride, was so discouraging that I have never been invited back. So far as I know they are still using miniature tests.

Reliability and Validity. Dr. Shimberg devoted considerable space to these concepts, and rightly so. As we know a test may be reliable and still not valid. The opposite is seldom if ever true.

Hopefully a reasonable degree of reliability is maintained in the California ratings of manipulative performance by following up the independent ratings obtained during the period of the test by a "jury" type rating. After the test is over the judges are asked to hold a "post mortem" for the purpose of comparing notes. They are requested to look for gross accidental errors as well as for deviances of more than two scale points. An adjusted score is obtained before the judges leave the test situation and while individual performance is still remembered. More often than not the three judges, representing labor, management, and the teaching profession find themselves within the stipulated two-point differential. Quite frequently, and to their pleasant surprise, they have assigned the identical rating score. The occasional judge who deviates widely and consistently from the other two is not invited to participate in the evaluation of subsequent tests.

The validity of the California manipulative tests is, so far, purely "internal" and consequently leans heavily upon the choice of the salient points selected for spot check. Reliable external criteria are notoriously difficult to establish. Inasmuch as a properly assembled work-sample test is virtually a criterion test, as Dr. Shimberg has pointed out, and since, for practical purposes, it seems "loves labor lost" to attempt getting closer to the criterion than the criterion itself I persist in recommending this type of synthetic test for the purpose at hand.

Rating Procedure. Preceding the test (frequently 12 to 15 different occupational tests are administered simultaneously on the same campus), instruction is given to the judges in the use of the linear-descriptive scale employed. The instruction also includes a brief overview of the meaning of reliability and validity and how to best preserve both and the importance of instant notation of the rating scores upon having formed an opinion, especially in rating procedures. Products are numbered and preserved for three months in case a dissatisfied candidate demands a review.

The rating scale used is a modification of The Harwood Industrial Efficiency Rating Scale appearing in a publication by H.C. Steinmetz, formerly chairman of the Department of Psychology, San Diego State College, entitled Manual of Industrial Efficiency Rating, 1943. Copies of the original and the modification are enclosed.

In closing I wish to say that, regardless of the image I may have created, my interest in experimental test research, and funds to permit it, is quite tremendous. My only plea is for sufficient theoretical research to go with it so as to prevent exact duplication of past errors in the practical application of manipulative proficiency tests.